

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 17:51:15 ; Search time 79.11 Seconds
(without alignments)
625.470 Million cell updates/sec

Title: US-09-265-585c-96
Perfect score: 3503
Sequence: 1 MPPPPPPPLTPYCRKCPPP.....KDLCLLTAASAMRPIQVPPCR 668

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101:*

- 1: /SID2/gcgdata/geneseq/AA1980.DAT.*
- 2: /SID2/gcgdata/geneseq/AA1981.DAT.*
- 3: /SID2/gcgdata/geneseq/AA1982.DAT.*
- 4: /SID2/gcgdata/geneseq/AA1983.DAT.*
- 5: /SID2/gcgdata/geneseq/AA1984.DAT.*
- 6: /SID2/gcgdata/geneseq/AA1985.DAT.*
- 7: /SID2/gcgdata/geneseq/AA1986.DAT.*
- 8: /SID2/gcgdata/geneseq/AA1987.DAT.*
- 9: /SID2/gcgdata/geneseq/AA1988.DAT.*
- 10: /SID2/gcgdata/geneseq/AA1989.DAT.*
- 11: /SID2/gcgdata/geneseq/AA1990.DAT.*
- 12: /SID2/gcgdata/geneseq/AA1991.DAT.*
- 13: /SID2/gcgdata/geneseq/AA1992.DAT.*
- 14: /SID2/gcgdata/geneseq/AA1993.DAT.*
- 15: /SID2/gcgdata/geneseq/AA1994.DAT.*
- 16: /SID2/gcgdata/geneseq/AA1995.DAT.*
- 17: /SID2/gcgdata/geneseq/AA1996.DAT.*
- 18: /SID2/gcgdata/geneseq/AA1997.DAT.*
- 19: /SID2/gcgdata/geneseq/AA1998.DAT.*
- 20: /SID2/gcgdata/geneseq/AA1999.DAT.*
- 21: /SID2/gcgdata/geneseq/AA2000.DAT.*
- 22: /SID2/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3503	100.0	668	21	AA28595	Maize Scarecrow pr
2	1861	53.1	653	18	AA28178	Arabidopsis SCAREC
3	1861	53.1	653	21	AA28569	Arabidopsis SCAREC
4	1131.5	32.3	716	18	AA28192	Maize ZCARECROM ZC
5	1131.5	32.3	716	21	AA28591	Maize ZCR protein.
6	1005.5	28.7	405	21	AA48516	Arabidopsis thalia
7	954.5	27.2	323	21	AA28603	Maize CBPPT4 ami
8	705.5	20.1	246	21	AA48517	Arabidopsis thalia
9	670	19.1	623	20	AA20240	Protein encoded by
10	666	19.0	630	20	AA20241	Protein encoded by
11	651	18.6	240	21	AA22966	Arabidopsis thalia

12	618	17.6	425	20	AA202539	Protein encoded by
13	594.5	17.0	587	22	AA202560	A. thaliana transc
14	594.5	17.0	587	22	AA201892	Arabidopsis thalia
15	590.5	16.9	517	21	AA38577	Arabidopsis thalia
16	590.5	16.9	518	21	AA38576	Arabidopsis thalia
17	590.5	16.9	532	18	AA30792	Arabidopsis thalia
18	590.5	16.9	533	21	AA38575	Arabidopsis thalia
19	589.5	16.8	531	21	AA28575	Arabidopsis thalia
20	589.5	16.8	532	22	AA202545	A. thaliana transc
21	589.5	16.8	532	22	AA201907	Arabidopsis thalia
22	589.5	16.8	572	22	AA31883	Amino acid sequenc
23	585.5	16.7	572	22	AA31884	Amino acid sequenc
24	548	15.6	384	21	AA329710	Arabidopsis thalia
25	548	15.6	405	21	AA29709	Arabidopsis thalia
26	548	15.6	413	21	AA29708	Arabidopsis thalia
27	544.5	15.5	480	21	AA21049	Arabidopsis thalia
28	544.5	15.5	578	21	AA21048	Arabidopsis thalia
29	544.5	15.5	593	21	AA21047	Arabidopsis thalia
30	532.5	15.2	587	21	AA28574	Arabidopsis thalia
31	525	15.0	428	19	AA261733	Tomato Ls protein.
32	502.5	14.3	306	21	AA28598	Protein encoded by
33	499.5	14.3	306	18	AA38179	Arabidopsis SCAREC
34	499.5	14.3	306	21	AA28570	Arabidopsis thalia
35	498	14.2	352	21	AA28596	Protein encoded by
36	492	14.0	352	18	AA28181	Arabidopsis SCAREC
37	466	13.3	541	21	AA28585	Arabidopsis thalia
38	466	13.3	543	21	AA28584	Arabidopsis thalia
39	466	13.3	610	21	AA28583	Arabidopsis thalia
40	439.5	12.5	513	21	AA28582	Arabidopsis thalia
41	439.5	12.5	515	21	AA28581	Arabidopsis thalia
42	439.5	12.5	583	21	AA28580	Arabidopsis thalia
43	423.5	12.1	375	21	AA28579	Arabidopsis thalia
44	404.5	11.5	287	21	AA28578	Arabidopsis thalia
45	404.5	11.5	809	21	AA28586	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AA28595	Arabidopsis standard; Protein: 668 AA.
ID	AA28595	
AC	AA28595	
DT	09-FEB-2001 (first entry)	
DE	Maize Scarecrow protein.	
XX	Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;	
KW	transgenic plant; cell division; molecular marker; herbicide resistance;	
KW	salt resistance; pathogen resistance; insect resistance.	
XX		
OS	Zea mays.	
XX		
PN	WO200053723-A2.	
XX		
PD	14-SEP-2000.	
XX		
PE	07-MAR-2000; 2000WO-US05875.	
XX		
PR	10-MAR-1999; 99US-0265585.	
XX		
XX	(UNYNY) UNIV NEW YORK STATE.	
PI	Bentley PN, Di Laurenzio L, Mysiocka-Diller J, Malamy JE, Pysch L,	
PI	Heierlitta Y, Bruce W, Lim J;	
XX		
DR	WPI; 2000-594315/56.	
XX	N-PSDB; AAC65299.	
PT	Scarecrow gene useful for producing transgenic plants expressing genes	
	whose product increases starch, lignin or cellulose biosynthesis and	

PT confers herbicide, pathogen or insect resistance -
 XX
 PS Claim 12; Fig 25; 200pp; English.
 XX
 CC The present sequence is given in a specification relating to the
 CC structure and function of a regulatory gene, SCARECROW (SCR).
 CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
 CC sequence similar to the sequence of MOTIF III (VHID) of Arabidopsis SCR
 CC protein are disclosed. SCR and SCL genes are useful for producing
 CC transgenic plants whose cell division is modified and root and/or stem
 CC development and gravitropism of stem or hypocotyl is altered. Cell
 CC division is increased or decreased in roots resulting in thicker or
 CC thinner root development. The transgenic plants are useful for expressing
 CC a gene of interest encoding a gene product that confers herbicide, salt,
 CC pathogen or insect resistance in root or embryo and genes encoding
 CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
 CC confers less susceptibility to lodging in the transgenic plants than a
 CC wild-type plant. SCR gene sequences are also useful as molecular markers
 CC for a quantitative trait e.g. root or gravitropism trait in molecular
 CC breeding of crop plants.
 CC
 XX
 SQ Sequence 668 AA;

Query Match 100.0%; Score 3503; DB 21; Length 668;
 Best Local Similarity 100.0%; Pred. No. 5,8e-269;
 Matches 668; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MPPPPPPPLPYCRCPRHPLPPPPSPNHLHYHOLDHOHAAAAAVRRKRPASDM 60
 Db 1 MPPPPPPPLPYCRCPRHPLPPPPSPNHLHYHOLDHOHAAAAAVRRKRPASDM 60
 OY 61 DLPPRRHVTGDLSDVTAAGVSGSPSSASQALPALTQHLQPLPAFOHHAPEVDY 120
 Db 61 dlpprrhvtgdlsgdvtaagvsgspssasqalpeltqlhqlpafghapevdy 120
 OY 121 PAHPRPAHAAOAGFATSTTAMVDGTRIDITIGSSGAASITQLIHNVRELIHPCNGL 180
 Db 121 pahprpahagagfatasttawvgllrdligssgaaavltqllhnvrelihpncpgl 180
 OY 181 ASLLELRSLRLAADPAPLPPPOFOHALHGAFAAPAGTLPPRPLPDKRHEHPP 240
 Db 181 asllelrslrlaadpaplppppqpqnaallhgapaapagtlclppppldokrtthpp 240
 OY 241 PCQOQOQOEHPHAPQSPKAPTAETATAAAAAAQAAMAAAKERKEEORRORDEGLTL 300
 Db 241 pcqggqgeehpqpapqspkaptaeetaaaaaaqaamaaakerkeegrrkdeeghl 300
 OY 301 TLLDCCARVANDNDIDAHQTLLETAELATPRTGSTORVAAVPAFAKARLVSSCLGIXA 360
 Db 301 tllldccarvandanndidahqtllelaelatrptgstgrvaayfaaamarsclgilya 360
 OY 361 PLPPGSPAARLHGKRVAAAFQVFNCSIPFVFSHTANQAIOEFEREERHIIIDLDIMQ 420
 Db 361 plppgspaarlhgvaaafqvfngisfpvfshftanqaiqeaefeerhiiidldimq 420
 OY 421 GLQWEGFLHILASRGGPPRVRLTGLGASMEALATYGRSLSDPDTLGLPEFCVAWEKA 480
 Db 421 glqwegflhilasrpgpprvrltglgsmealeatgrkslfdadtlglpfcavaeka 480
 OY 481 GNVDEKLGVTTRREAVAVHMLHSLYDVTGSDSNLMLIQRLAKRVYVMWDJLSHSGSF 540
 Db 481 gnvdekigvttrreavavhmlhsltydvtsdntlmliqlrakrvyvmwedjshsgsf 540
 OY 541 LARFEALHYYSALFDSDASVGEDSPRHVVEOQLSRETRNVLAVGAPARTGDVKGFS 600
 Db 541 larfealhyysalfdsdasvgedsperhveeqllsretrnvlayvgparfydvkfgs 600
 OY 601 WRKLAQSGFRAASLAGSAAAQASLLQMFPSDGYTTLVEENGAKLKGKDCILTLTASMR 660
 Db 601 wrkлагsfraaslagasaaqasllqgmfsdgytlveengalklgykdciltaaawr 660
 OY 661 PIQVPPCR 668

Db 661 piqvppcr 668

RESULT 2
 AAW38178
 ID AAW38178 standard; Protein; 653 AA.
 XX
 AC AAW38178;
 XX
 DT 11-MAY-1998 (first entry)
 XX
 DE Arabidopsis SCARECROW protein.
 XX
 KW SCARECROW; SCR gene; transgenic plant; root; gravitropism;
 KW crop improvement.
 XX
 OS Arabidopsis thaliana.
 OS
 XX
 FH Key
 FH Region
 FT Location/Qualifiers
 FT 1..267
 FT /note= "N-terminal region rich in Gln, Ser, Thr and
 FT Pro homopolymeric structures"
 FT Peptide
 FT 291..322
 FT /note= "Leucine heptad repeat"
 FT Peptide
 FT 436..473
 FT /note= "Leucine heptad repeat"
 FT Domain
 FT 58..99
 FT /label= MOTIF-I
 FT Domain
 FT 265..341
 FT /label= MOTIF-II
 FT Domain
 FT 373..435
 FT /label= MOTIF-III (VHID)
 FT Domain
 FT 436..480
 FT /label= MOTIF-IV
 FT Domain
 FT 457..496
 FT /label= MOTIF-V
 FT Domain
 FT 516..649
 FT /label= MOTIF-VI
 XX
 PN W09741152-A1.
 XX
 PD 06-NOV-1997.
 XX
 XX 25-APR-1997; 97MO-US07022.
 PF
 XX 24-APR-1997; 97US-0842445.
 PR 26-APR-1996; 96US-0638617.
 XX
 XX (UYNX) UNTV NEW YORK STATE.
 PA
 XX
 PI Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JE;
 PI Pysl L, Myssocka-Diller J;
 XX
 DR WPI; 1997-549683/50.
 DR N-PSDB; AAT95753.
 XX
 PT DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
 PT cell division and therefore alter root development, or alter plant
 PT stem or hypocotyl gravitropism
 XX
 PS Claim 12; Page 92-94; 221pp; English.
 XX
 CC This is the deduced amino acid sequence of the Arabidopsis
 CC SCARECROW (SCR) gene (see AAT95753). SCR is a member of a novel
 CC protein family and contains a number of potential functional
 CC domains similar to those found in transcription factors. SCR is
 CC expressed specifically in embryo root progenitor tissue and in
 CC certain root and stem tissues. It regulates a specific asymmetric
 CC division, and controls gravitropic response in aerial structures
 CC and root formation. SCR proteins (see AAW38178-201) from dicots and
 CC monocots, SCR proteins lacking 1-4 of MOTIFS I-VI, and polypeptides
 CC corresponding to MOTIF I, II, III, IV, V or VI of SCR are claimed.

Transgenic plants can be engineered to overexpress the SCR protein, so that cell division is increased in roots, resulting in thicker root development, while a plant with an altered stem or hypocotyl gravitropism is less susceptible to lodging. Plants that contain an antisense molecule that suppresses the expression of endogenous SCR gene product have thinner root development. A gene of interest can be placed under control of a SCR promoter and expressed in a plant to confer herbicide, salt, pathogen or insect resistance, or when expressed in stems to increase starch, lignin or cellulose biosynthesis (all claimed).

Sequence 653 AA:

Query Match 53.1%; Score 1861; DB 18; Length 653;
Best Local Similarity 57.3%; Pred. No. 6, 2e-139;
Matches 404; Conservative 66; Mismatches 123; Indels 112; Gaps 18;

4 PPPPPPL-----TPYCRRCPPPHLPPPPSSPNHFLHLHLDHQAAMAAWVKRP 56
12 PPPHSPILRTSSGSSSSNNRPPP-PPPP-----LVMVTKRL 48
57 ASDM-----DLPPRR--H-VTGDLSDVT-----AAAAGVGGSGAPS---SASAQ 96
49 asemssnpdymnsrrprvshlidsnyntvtpqgppltaaatvssqpnplsvcgfsq 108
97 LPLPTQLHQLPRAFHNAPEVDVPAHPAPAAHQAAGCATASTTAWDGIIRDIIGSSG 156
109 lpfvfrs-----drggrtnvmsvqpmnd---qdssssasptvwdaifrdlhlss- 154
157 GAASITQLIHNVRELIHPCNPGLASILELRSLAADPA---PLP----- 200
155 -tsvisipqlhgvrdlflfcpnplgallelyrlrslmldpsssdpspqtfeptyqism 213
201 --PPPOQHALLHGAPAAAPAGTLTLPPLPDKRRHHPCCQQOQOQEEHPAPQSPK 258
214 pspqgqgqbq-----gqgqbhkprrppidqgere-----nstd 248
259 APYAEETAAAA-AAQAAAAAAKERKEEQRKORDEBGLHLLTLTLCQAEVADNMLDD 317
249 appqpetvatvpavqntlaealrkrkeelkrqdeegllhlllllqcaevadnlee 308
318 AHQTLLELAELATPGCTSTORVAVPAEAMASRLVSSCLGYALPFGSPRAAALHG-RV 376
309 anklllelsglspqysaqvvaayfseamaarlinsclgylaLP--strwmpqthslkm 366
377 AAARQVFNIGIPFYKFESHFTANQAIQAEFEREREVHIIIDIMQGLOMPGLFHLIASRPG 436
367 vsatqvfngiseplykfstftangaiqeafekedsvhllldimglqwpqlfhljastrpy 426
437 GPPRVRLTGASMEALEATGKRLSDPADTLGLPFECAVAEKAGNVDEPKLGVTRREAV 496
427 gpphvrllgltlsmeaiaqatgkrlsdfdklglpfeicpplaekvgnldterlnvrkreav 486
497 AVHMLHSLVYVNTSGSDSTMLLQRLAPKYVTWYEOPLSHSGSLAFVFAIHVYSALFD 556
487 avhmlqshlyvntsgsdstmlwlqrlapkyvtlvveqdlshagsflgrtveahlyysalfd 546
557 SLDASGDESPERHVEEQQLSRIRNVLAVAGPARTGDVFGSWRKRKLAKOSGRASLAS 616
547 slgasygeeserhveeqqlskelrnlavagpsrsgervkfweswrekmqcgfkglsla 606
617 GSAAAQSLILGMPFSDGYTLVEENGALIKGWDKLCILTLASAMRP 661
607 gnaatqatalligmfrsdgyltlvdnglklgwkqlsiltasawrp 651

RESULT 3
AAB28569
ID AAB28569 standard; Protein; 653 AA.
XX
AC AAB28569;
XX

DT 09-FEB-2001 (first entry)
XX Arabidopsis SCARECROW protein.
DE
XX Scarecrow gene: SCR gene; plant: regulatory gene; SCARECROW-like; SCL;
KW transgenic plant; cell division; molecular marker; herbicide resistance;
XX salt resistance; pathogen resistance; insect resistance.
XX
OS Arabidopsis thaliana.
XX
PN WO200053723-A2.
XX
PD 14-SEP-2000.
XX
PF 07-MAR-2000; 2000WO-US05875.
XX
PR 10-MAR-1999; 99US-0265585.
XX
PA (UYNT) UNIV NEW YORK STATE.
XX
XX Benfey PN, Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysch L;
PI Helariutta Y, Bruce W, Lim J;
XX
DR MPI: 2000-594315/56.
XX
XX N-PSDB; AAC65280.
XX
PT Scarecrow gene useful for producing transgenic plants expressing genes
PT whose product increases starch, lignin or cellulose biosynthesis and
XX confers herbicide, pathogen or insect resistance -
XX
PS Claim 2; Fig 5A; 200pp; English.

The present sequence is given in a specification relating to the structure and function of a regulatory gene, SCARECROW (SCR). SCARECROW-like (SCL) genes encoding proteins containing an amino acid sequence similar to the sequence of MOTIF III (WHIRL) of Arabidopsis SCR protein are disclosed. SCR and SCL genes are useful for producing transgenic plants whose cell division is modified and root and/or stem development and gravitropism of stem or hypocotyl is altered. Cell division is increased or decreased in roots resulting in thicker or thinner root development. The transgenic plants are useful for expressing a gene of interest encoding a gene product that confers herbicide, salt, pathogen or insect resistance in root or embryos and genes encoding starch, lignin or cellulose biosynthesis in shoots. The SCR gene also confers less susceptibility to lodging in the transgenic plants than a wild-type plant. SCR gene sequences are also useful as molecular markers for a quantitative trait e.g. root or gravitropism trait in molecular breeding of crop plants.

Sequence 653 AA:

Query Match 53.1%; Score 1861; DB 21; Length 653;
Best Local Similarity 57.3%; Pred. No. 6, 2e-139;
Matches 404; Conservative 66; Mismatches 123; Indels 112; Gaps 18;

4 PPPPPPL-----TPYCRRCPPPHLPPPPSSPNHFLHLHLDHQAAMAAWVKRP 56
12 PPPHSPILRTSSGSSSSNNRPPP-PPPP-----LVMVTKRL 48
57 ASDM-----DLPPRR--H-VTGDLSDVT-----AAAAGVGGSGAPS---SASAQ 96
49 asemssnpdymnsrrprvshlidsnyntvtpqgppltaaatvssqpnplsvcgfsq 108
97 LPLPTQLHQLPRAFHNAPEVDVPAHPAPAAHQAAGCATASTTAWDGIIRDIIGSSG 156
109 lpfvfrs-----drggrtnvmsvqpmnd---qdssssasptvwdaifrdlhlss- 154
157 GAASITQLIHNVRELIHPCNPGLASILELRSLAADPA---PLP----- 200
155 -tsvisipqlhgvrdlflfcpnplgallelyrlrslmldpsssdpspqtfeptyqism 213
201 --PPPOQHALLHGAPAAAPAGTLTLPPLPDKRRHHPCCQQOQOQEEHPAPQSPK 258

```

Db 214 pspqqqqqhq-----:qgqqqhpppppqqgqere-----nsstd 248
QY 259 APTAEFTAAAA-AAQAAAAAAKKEKQRRKQDEBGLHLLTLLOCAEVAMDND 317
Db 249 appqpetvatvavqtnlaaalrerekirkqkqdeqhlhlllllqcaevsadnl 308
QY 318 AHQTLLEIAELATPECTSTQRYAAVFAEAMSAFLVSSCIGLAPPCGPAARLHG-RV 376
Db 309 ankllleisqlstpytsaqryaayfseamsarlllncscgllyaalp--symmpqchslkm 366
QY 377 AAAFVFNCGISPPVKFSHTANQAIQEAEREERVHIIDLDIMGLQWFGLEPHILASRPG 436
Db 367 vsaifgvingisplvktfshltanqalqeaekedsvhlldldimgllywpglfhllasrpg 426
QY 437 GPPRVLTGLGASMEALTEGKRSLSPFADTLGPRFCAVAKANVDEKIGVRRRAV 496
Db 427 gpphvtlgltsmealgatgkrlsdfctoklglpfctoplaekvgnldehlnvkrav 486
QY 497 AVHMLHSLYDVYDSDSNLTAMLQRLAPKVTVMVEDLSHSGSFLARFEVAIHYSALFD 556
Db 487 avhwlgshlydvrtgsdahnltwlglrlapkvvtlvveqdlshagsflgrtveahlysalfd 546
QY 557 SLDASTGEDSPERHVVBOQLSRETRNVILAVGPARTDVKGFGVREKLAQSGFPAASLA 616
Db 547 slgasysgeeseehrhveqqlskeltrnvilavgparsgvktfsvrekmqcgfkqisla 606
QY 617 GSAAGASLLGMPSPDGYTLWEENGALKLGKDCLLLTASWMP 661
Db 607 gnaatqatclllgmfpdgytlvdhngclklgwkdslilasawcp 651

```

RESULT 4

ID AAM38192 standard: Protein; 716 AA.

AC AAM38192;

DT 11-MAY-1998 (first entry)

XX Maize ZCARECROW ZCR protein.

XX ZCARECROW; ZCR; SCARECROW; SCR; transgenic plant; root;

KW gravitropism; crop improvement; maize.

XX Zea mays.

XX Key Location/Qualifiers

```

FT 7 Misc-difference 7 /note= "encoded by TAA"
FT 9 Misc-difference 9 /note= "encoded by TAA"
FT 15 Misc-difference 15 /note= "encoded by TGA"
FT 19 Misc-difference 19 /note= "encoded by TGA"
FT 20 Misc-difference 20 /note= "encoded by TGA"
FT 26 Misc-difference 26 /note= "encoded by TGA"
FT 31 Misc-difference 31 /note= "encoded by TGA"
FT 34 Misc-difference 34 /note= "encoded by TGA"
FT 35 Misc-difference 35 /note= "encoded by TGA"
FT 38 Misc-difference 38 /note= "encoded by TGA"
FT 46 Misc-difference 46 /note= "encoded by TGA"
FT 48 Misc-difference 48 /note= "encoded by TGA"
FT 62 Misc-difference 62 /note= "encoded by TGA"

```

```

FT 65 /note= "encoded by TGA"
FT 68 /note= "encoded by TAG"
FT 82 /note= "encoded by TAG"
FT 113 /note= "encoded by TAG"
FT 114 /note= "encoded by TGA"
FT 134 /note= "encoded by TAG"
FT 136 /note= "encoded by TGA"
FT 148 /note= "encoded by TGA"
FT 184 /note= "encoded by TAA"
FT 188 /note= "encoded by TAA"
FT 190 /note= "encoded by TAG"
FT 239 /note= "encoded by TAG"
FT 245 /note= "encoded by TAG"
FT 259 /note= "encoded by TGA"
FT 316 /note= "encoded by TAA"
FT 318 /note= "encoded by TGA"
FT 378 /note= "encoded by TGA"
FT 386 /note= "encoded by TAG"
FT 586 /note= "encoded by TAG"
FT 615 /note= "encoded by TAA"
FT 625 /note= "encoded by TGA"
FT 669 /note= "encoded by TGA"
FT 671 /note= "encoded by TAG"
FT 674 /note= "encoded by TAA"
FT 694 /note= "encoded by TGA"
FT 697 /note= "encoded by TAG"
FT 707 /note= "encoded by TAG"
FT 318..539 /note= "encoded by TGA"
FT /note= "region of homology to SCR"

WO9741152-A1.
PN
XX
XX 06-NOV-1997.
PD
XX
XX 25-APR-1997; 97MO-US07022.
PF
XX 24-APR-1997; 97US-0842445.
PR
XX 26-APR-1996; 96US-0638617.
XX
XX (UYNY ) UNIV NEW YORK STATE.
PA
XX Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JE;
PI Pysh L, Wyszocka-Diller J;
XX
XX WPI: 1997-549683/50.
DR N-PSDB; AAT95767.
XX
XX DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
XX cell division and therefore alter root development, or alter plant

```


PT stem or hypocotyl gravitropism
 XX
 PS Claim 2; Page 138-140; 221pp; English.
 XX
 CC This is the deduced amino acid sequence of the maize ZCARECROW gene
 CC (see AAT95767). It has a region of homology to Arabidopsis SCARECROW
 CC (SCR) protein (see AAM38178). SCR is a member of a novel protein
 CC family and contains a number of potential functional domains
 CC similar to those found in transcription factors. SCR is expressed
 CC specifically in embryo root progenitor tissue and in certain root
 CC and stem tissues. It regulates a specific asymmetric division, and
 CC controls gravitropic response in aerial structures and root
 CC formation. SCR proteins (see AAM38178-201) from dicots and monocots,
 CC corresponding to MOTIF I, II, III, IV, V or VI of SCR are claimed.
 CC Transgenic plants can be engineered to overexpress the SCR protein,
 CC so that cell division is increased in roots, resulting in thicker
 CC root development, while a plant with an altered stem or hypocotyl
 CC gravitropism is less susceptible to lodging. Plants that contain
 CC an antisense molecule that suppresses the expression of endogenous
 CC SCR gene product have thinner root development. A gene of interest
 CC can be placed under control of a SCR promoter and expressed in a
 CC plant to confer herbicide, salt, pathogen or insect resistance, or
 CC when expressed in stems to increase starch, lignin or cellulose
 CC biosynthesis (all claimed).
 CC
 SQ Sequence 716 AA:
 Query Match 32.3%; Score 1131.5; DB 18; Length 716;
 Best Local Similarity 78.1%; Pred. No. 4.2e-81;
 Matches 225; Conservative 12; Mismatches 24; Indels 27; Gaps 3;
 QY 234 RREHPPPCQQOQDEPHAPQSPKAPTAETAAAAAQAQAAAAAAKERKEBQRKORD 293
 Db 279 rrrsrp-----rrpprprrpprr--sgrtcsygssatrrastcx----- 316
 QY 294 EEGHLLTLTLCCAEVNAINDDAHOTLLEIAELATPFGTSTORVAAYPAEAMSARLV 353
 Db 317 -----vllllqcaevnadnldahqtllelaelalpigtscqvaayfaeamsarlv 371
 QY 354 SCLGLYAPLPSPSPAARLHGRVAAAFQVFNQISPFVKFSHTANQAIOEAFREERYHI 413
 Db 372 sclgilyaplpsspsaarllngryaaafqvfngispfvkfshtanqalgeafereerh 431
 QY 414 IDLDIMQGLQMPGLFHILASRPGSPRRVRLTGLGASMEALEATGKRISDPADTLGLPEF 473
 Db 432 idldimglqmpglfhlilvsrpgsprvrltglgsmdaleatgkrisdfadtlglpfe 491
 QY 474 CAVAEKAGNDPEKLGTVRREAVAVHMLHSLDYDVGSDSNTIMLQIR 521
 Db 492 cavaekagnvdpqklgytrreavavhphnslidyvsgdsntlwlilqr 539
 RESULT 5
 AAB28591 standard; Protein; 716 AA.
 ID AAB28591
 AC AAB28591;
 XX
 DT 09-FEB-2001 (first entry)
 XX
 DE Maize ZCR protein.
 XX
 KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
 KM transgenic plant; cell division; molecular marker; herbicide resistance;
 KW salt resistance; pathogen resistance; insect resistance.
 XX
 OS Zea mays.
 XX
 PN MO200053723-A2.
 XX
 PD 14-SEP-2000.

XX
 PF 07-MAR-2000; 2000MO-US05875.
 XX
 PR 10-MAR-1999; 99US-0265585.
 XX
 PA (UWNY) UNIV NEW YORK STATE.
 XX
 PI Benfey PN, Di Laurenzio L, Wyszoka-Diller J, Malamy JE, Pysh L;
 PI Helariutta Y, Bruce W, Lim J;
 DR WPI; 2000-594315/56.
 DR N-PSDB; AAC65298.
 XX
 PT Scarecrow gene useful for producing transgenic plants expressing genes
 PT whose product increases starch, lignin or cellulose biosynthesis and
 PT confers herbicide, pathogen or insect resistance
 CC
 CC Claim 2; Fig 17B; 200pp; English.
 CC
 CC The present sequence is given in a specification relating to the
 CC structure and function of a regulatory gene, SCARECROW (SCR).
 CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
 CC sequence similar to the sequence of MOTIF III (VHIII) of Arabidopsis SCR
 CC protein are disclosed. SCR and SCL genes are useful for producing
 CC transgenic plants whose cell division is modified and root and/or stem
 CC development and gravitropism of stem or hypocotyl is altered. Cell
 CC division is increased or decreased in roots resulting in thicker or
 CC thinner root development. The transgenic plants are useful for expressing
 CC a gene of interest encoding a gene product that confers herbicide, salt,
 CC pathogen or insect resistance in root or embryo and genes encoding
 CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
 CC confers less susceptibility to lodging in the transgenic plants than a
 CC wild-type plant. SCR gene sequences are also useful as molecular markers
 CC for a quantitative trait e.g. root or gravitropism trait in molecular
 CC breeding of crop plants.
 CC
 SQ Sequence 716 AA:
 Query Match 32.3%; Score 1131.5; DB 21; Length 716;
 Best Local Similarity 78.1%; Pred. No. 4.2e-81;
 Matches 225; Conservative 12; Mismatches 24; Indels 27; Gaps 3;
 QY 234 RREHPPPCQQOQDEPHAPQSPKAPTAETAAAAAQAQAAAAAAKERKEBQRKORD 293
 Db 279 rrrsrp-----rrpprprrpprr--sgrtcsygssatrrastcx----- 316
 QY 294 EEGHLLTLTLCCAEVNAINDDAHOTLLEIAELATPFGTSTORVAAYPAEAMSARLV 353
 Db 317 -----vllllqcaevnadnldahqtllelaelalpigtscqvaayfaeamsarlv 371
 QY 354 SCLGLYAPLPSPSPAARLHGRVAAAFQVFNQISPFVKFSHTANQAIOEAFREERYHI 413
 Db 372 sclgilyaplpsspsaarllngryaaafqvfngispfvkfshtanqalgeafereerh 431
 QY 414 IDLDIMQGLQMPGLFHILASRPGSPRRVRLTGLGASMEALEATGKRISDPADTLGLPEF 473
 Db 432 idldimglqmpglfhlilvsrpgsprvrltglgsmdaleatgkrisdfadtlglpfe 491
 QY 474 CAVAEKAGNDPEKLGTVRREAVAVHMLHSLDYDVGSDSNTIMLQIR 521
 Db 492 cavaekagnvdpqklgytrreavavhphnslidyvsgdsntlwlilqr 539
 RESULT 6
 AAG48516 standard; Protein; 405 AA.
 ID AAG48516
 AC AAG48516;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 61274.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135144.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147403.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

```

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 28.7%; Score 1005.5; DB 21; Length 405;
 Best Local Similarity 53.2%; Pred. No. 1.9e-71;
 Matches 207; Conservative 65; Mismatches 106; Indels 11; Gaps 6;

```

QY 278 AAEERKE--EORRKORDEGILHTLTLLOCAEVAENDNDHQTLLLETAELATPGTS 335
      :| | | | | : : :| | | | | | | | | | | | | | | | | | | | | | |
DB 17 ssaakrifepeetlendgaaaiiklslilqcaeyavachlreaatlsetseicspgss 76
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
QY 336 TORVAAYFAEAMASRLVSSCT-GLYAPLPQSPAAARLHGRVAARFOVNGISPFVKFSH 394
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB 77 perwayfaagqlgrvlsyisgacpslse-kplttvgsqkifaealqlysvaplkfsh 135
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
QY 395 FTAAOAIQAEERERERHIIIDLMQGLQWPGFLHILASRPGGPPRRRLTGLGASKEALE 454
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB 136 ftanqaifgaIdgedsvhlldldvmgqlwpalfhllasprklrsrltfgssdlla 195
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
QY 455 ATGRRLSDPADTGLPFEFCVAEKAGN-VDPEKLGVTREAVAVHMLHHSILYDVGTSDS 513
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB 196 stgrtladfaaslnlpfeifpnieiglnldpsqlatrgaeavvhwqhrlYdvtgnml 255
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
QY 514 NTLMLEIQLARVVTWVEQDLSH--SGSFLARFEALHYYSALFSDLSADYSGEDSPERHV 571
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
DB 256 etleilrlkplnltlvvegelsyddgsflgrfvealhyysalfdaIdgdglgeesgrftc 315
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |
QY 572 VEQQLLSREIRNVLAVGVPARTGVKFGSWREKLASGCFRAASLAGSAAQASILLGMFP 631
      :| | | | | :| | | | | | | | | | | | | | | | | | | | | | |

```

```

DB 316 vegivlgteirlnlvahggrr-----krmkweelsrvgrfsvlrgnpatgqililmpl 371
QY 632 SDGYTLVEENGALKLGWKDCLLTATASMR 660
      :| | | | | | | | | | | | | | | | | | | | | | |
DB 372 wngylveenglrlrgwkdslltsaaw 400

```

RESULT 7
 AAB28603
 ID AAB28603 standard; Protein; 323 AA.
 XX
 AC AAB28603;
 XX
 DT 09-FEB-2001 (first entry)
 XX
 DE Maize CBPRT44 amino acid sequence.
 XX
 KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
 KW transgenic plant; cell division; molecular marker; herbicide resistance;
 KW salt resistance; pathogen resistance; insect resistance.
 XX
 OS Zea mays.
 XX
 PN W0200053723-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 07-MAR-2000; 2000MO-US05875.
 XX
 PR 10-MAR-1999; 99US-0265585.
 XX
 PA (UYNY) UNIV NEW YORK STATE.
 XX
 PI Bentley PN, Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pyeh L,
 PI Helariutta Y, Bruce W, Lim J;
 DR N-PSDB; AAC65334.
 XX
 WPI; 2000-594315/56.

PT Scarecrow gene useful for producing transgenic plants expressing genes
 PT whose product increases starch, lignin or cellulose biosynthesis and
 PT confers herbicide, pathogen or insect resistance
 XX
 PS Example 8; Fig 33; 2000p; English.
 XX
 CC The present sequence is given in a specification relating to the
 CC structure and function of a regulatory gene, SCARECROW (SCR).
 CC SCARECROW-like (SCL) genes encoding proteins containing an amino acid
 CC sequence similar to the sequence of MYO1F III (VH1D) of Arabidopsis SCR
 CC protein are disclosed. SCR and SCL genes are useful for producing
 CC transgenic plants whose cell division is modified and root and/or stem
 CC development and gravitropism of stem or hypocotyl is altered. Cell
 CC division is increased or decreased in roots resulting in thicker or
 CC thinner root development. The transgenic plants are useful for expressing
 CC a gene of interest encoding a gene product that confers herbicide, salt,
 CC pathogen or insect resistance in root or embryos and genes encoding
 CC starch, lignin or cellulose biosynthesis in shoots. The SCR gene also
 CC confers less susceptibility to lodging in the transgenic plants than a
 CC wild-type plant. SCR gene sequences are also useful as molecular markers
 CC for a quantitative trait e.g. root or gravitropism trait in molecular
 CC breeding of crop plants.
 CC
 SO Sequence 323 AA.

Query Match 27.2%; Score 954.5; DB 21; Length 323;
 Best Local Similarity 62.9%; Pred. No. 1.5e-67;
 Matches 190; Conservative 37; Mismatches 66; Indels 9; Gaps 3;
 QY 369 AARLHGRVAARFOVNGISPFVKFSHTANQAIQAEERERERHIIIDLMQGLQWPGFLF 428
 DB 1 aaagstrvavafqaynalspvksfhtangalldgdcdhvidldimglqwp9l 60


```

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147203.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149380.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160774.
PR 21-OCT-1999; 99US-0160777.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

```

```

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 20.1%; Score 705.5; DB 21; Length 246;
Best Local Similarity 57.6%; Pred. No. 5.5e-48;
Matches 141; Conservative 37; Mismatches 60; Indels 7; Gaps 3;

```

```

QY 419 MOGLQWPGFLHTLASRPGPPRRVRLTGLASMEALFATGKRSDPADTGLPFEECAVAE 478
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 mgq1gwpa1fnh1asrpk1re1rt1gtf1gsscd1lastr1ad1f1ass1n1pf1ehpleg 60
QY 479 KAGN-VDPEKLGTVRREAVAVHMLHSLVDYDSDSNTLMLQLR1APKVVYTWEDLSH- 536
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ilgn1ldpsqlatrg1eeavvhw1mghr1ydv1cgm1tel1el1r1k1n1l1t1ve1ge1jd 120
QY 537 -SGSFLARFVEA1HYSA1FDSD1DAS1YGDS1PBR1HYVEQ1LLS1RE1RNV1LAV1G1P1RTGD 595
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 dgs1flgr1tvea1lhy1sal1fd1l1gd1l1ge1sger1tve1g1v1g1el1n1v1ah1g99rr-- 177
QY 596 VKFGSWREK1LAGSFRRA1LAGSAAQAASLL1GMFP1SDGYTV1EENGAL1K1GK1D1CL1LT 655
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 -k1mkwkeel1strv1g1r1pval1r1gnpat1gag1ll1gm1pwm1gyl1veeng1rl1gwk1d1s1lt 236
QY 656 ASAMR 660
  |||||
Db 237 asaw1k 241

```

```

RESULT 9
ID AAY02540
AA02540 standard; Protein; 623 AA.
XX
AC AAY02540;
XX
DE 16-JUL-1999 (first entry)
XX
DE Protein encoded by wheat Rht clone 5a1 genomic sequence.
XX
DE Rht gene: homologue; Triticum aestivum; wheat; growth inhibition;
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW paclobutrazol.
XX
OS Triticum aestivum.
XX
PN WO9909174-A1.
XX
PD 25-FEB-1999.
XX
PF 07-AUG-1998; 98WO-GB02383.
XX
PR 13-AUG-1997; 97GB-0017192.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
PI Harberd NP, Peng J, Richards DE;
XX
DR WPI: 1999-181040/15.
DR N-PSDB; AAX36279.
XX
XX New Triticum aestivum polynucleotides - encode a polypeptide which
PT provides inhibition of the growth of plants, which inhibition is
PT antagonised by gibberellin, used to confer a dwarf phenotype
XX
XX Disclosure; Fig 8p; 88pp; English.
XX
XX The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in Triticum aestivum, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable

```

by treatment with gibberellin. In addition, the products can be used to produce Rht mutant plants which are dwarfed compared with wild-type, the dwarfing being gibberellin-insensitive. Taller plants may be made by knocking out Rht or the relevant homologous gene in the plant of interest. Plants may be made which are resistant to compounds which inhibit gibberellin biosynthesis, such as paclobutrazol, e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds dwarf but let crop plants grow tall. The present sequence is encoded by the wheat Rht clone 5a1 genomic sequence.

Sequence 623 AA:

Query Match 19.1%; Score 670; DB 20: Length 623;

Best Local Similarity 30.9%; Pred. No. 1,2e-44; Matches 202; Conservative 77; Mismatches 262; Indels 112; Gaps 19;

```

OY 79 AAAGVGGSGAPSSASAOPLALPTQLHQLPRAFGHNAPEVDVPAHPARAHAQAGGATA 138
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 8 aagsggsggsmgsgedkmmvs-----aaagegeevd-----ellaalykvr 50

OY 139 SRTAVVDGIIRDI-----IGSSG---GAAYSIQILHNAREIHPNPGIASLLELRSL 192
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 51 scmadvaqkleglemamgmgygagapddsfathtatctvhyprtdssvvesml-sel 109

OY 193 AADPARLPDPQ-----PQHALLHGAPAAAPAGLTLP---PP 227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 110 napprlppaglnaststvgsgyfdlppsvdssslvalpripapatacpadisa 169

OY 228 PPLPRKRHNPPRCQQQQQDEPPHAPDSPAFAEETAAAAAQAQAAAAAERKEEQ 287
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 170 dsvirpkrmtlrgsstssssssslggagavveaarpvaaaanlprlvv----- 224

OY 288 RRKQDEGLHLTLTLCQAEVNVNDNDHQTLETAELATPEFGSTOKAAYFAEAM 347
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 225 --vdtgeagirlvhaliceaeavgenlsaaeklykqrlpilaasggamtkvaaylgeal 282

OY 348 SARLVSSCLGLYAPLPGS---PAARLHGRVAAAFQVNGISPEVYKFSHTANQAOIE 403
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 283 arvrfr-----frppdslldaafadllhab-----fyescpylkkfhtangalle 330

OY 404 AFERERVHIIDLMQGLMOPGFLHILASRPGPRVRLTGCA---SKEALEATGKR 459
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 331 atagcrrvhyvdfgikgmpwpaillqalalrpgppsfllrpgppdcdalqygvw 390

OY 460 LSDFADTLGLPEEFCVAV-----EKAGNVDPKRGVAREAVAHMLH--HSL 505
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 391 laqfahltlrvtfdqyrglvaatladerfmigpegeedpne-----epeviahsvfemhrl 446

OY 506 YDVTGSDSNTLMLORLAPKVVTVWEDLSH--SGSFARFVEALHYXSALEDSLDASYGE 564
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 447 laqpalalekltgvtvaavprlvtvgeaahnsqtfldrfeslhysltmdstlegssg 506

OY 565 DSPER-----HYVEQQLSREIRNVAVGGAAR--GDVKGSGWKEKLAQS 608
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 507 ggpsevsqgaapaagatdgvmevylygdlcnvacegegeterhelcigwvtnrigna 566

OY 609 GFRAASLAGSAAQAASLIGMFP--SDGYTLVEENGALKLGKDCILTLTASAMR 660
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 567 gfevthlgsnakyqastllalfagagdykveekegoltlgtvhttrpllataswr 619

RESULT 10
AA02541
ID AA02541 standard; Protein: 630 AA.
XX
XX AA02541:
XX
XX 16-JUL-1999 (first entry)
XX
XX Protein encoded by maize 1a1 genomic clone sequence.
DE
XX
XX Rht gene; homologue: Triticum aestivum; wheat; growth inhibition;

```

KW antagonist: gibberellin; dwarf phenotype; gibberellin biosynthesis;
 XX paclobutrazol; maize.
 OS
 XX Zea mays.
 PN MO9909174-A1.
 XX
 PD 25-FEB-1999.
 XX
 PF 07-AUG-1998; 98WO-GB02383.
 XX
 PR 13-AUG-1997; 97GB-0017192.
 XX
 PA (PLAN-) PLANT BIOSCIENCE LTD.
 XX
 PI Harberd NP, Peng J, Richards DE;
 XX
 DR WPT; 1999-181040/15.
 XX N-PSDB; AAX36280.

PT New Triticum Aestivum polynucleotides - encode a polypeptide which
 PT provides inhibition of the growth of plants, which inhibition is
 PT antagonised by gibberellin, used to confer a dwarf phenotype
 PS
 PS Disclosure; Fig 9b; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and
 CC its homologues) that, when expressed in Triticum Aestivum, inhibit
 CC growth of the plant. This growth inhibition is antagonised by
 CC gibberellin. The products can be used to provide Rht expression in
 CC plants, conferring a dwarf phenotype on a plant which is correctable
 CC by treatment with gibberellin. In addition, the products can be
 CC used to produce Rht mutant plants which are dwarfed compared with
 CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
 CC may be made by knocking out Rht or the relevant homologous gene in
 CC the plant of interest. Plants may be made which are resistant to
 CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol,
 CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
 CC dwarf but let crop plants grow tall. The present sequence is encoded by
 CC the maize 1a1 genomic clone sequence.

Sequence 630 AA:

Query Match 19.0%; Score 666; DB 20: Length 630;

Best Local Similarity 31.9%; Pred. No. 2.6e-44; Matches 209; Conservative 80; Mismatches 253; Indels 114; Gaps 21;

```

OY 71 GDLSDVTAANAAGVGGSGAPSSASAOPLALPTQLHQLPRAFGHNAPEVDVPAHPARAHA 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 16 gskdkmmaaaagag-----eqeedvd-----ella 42

OY 131 QACGEATASTAVMDGIIRDI-----IGSSGAANSITO--LIHNAREIHPNPGIASL 183
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 43 algykvrssdmadvakleglemamgmgygagatdgvfshlatdvhypsdlsaw 102

OY 184 LELRLSLLAADPARLP--PPOQHALLHGAPAAAPAGLTLP---PPRDLKRRNH--- 238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 219 gasaaanpavrvvvdvtg--eaqrlvhaliceaeavgenlfisaaalvqxjplmlassg 276

OY 333 GTSTORVAAVFAAMSARLVSSCLGLYAPLPSPAAARLHGRVAAAFQV--FNGISPEVK 391
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 277 ggamtkvaaylgealarr-----vytrfrp--pdslldaafadllhanfyescpylk 327

OY 392 FSHFTANQAOIEAFERERVHIIDLMQGLMOPGFLHILASRPGPRVRLTGIGA--- 448

```

```
Db      328 fahntaagallleaifagcrrrvhvdfikgmgwpaallgaalrpgspfrltqvgppp 387
Qy      449 -SMEALFATGKRSLSDPADTLGLDFEECA-VAEKAGVDEPKLGV-----TREAVAVIWM 500
Db      388 decdalqvgwklagfahntlrvdqfgyrlvaatladleptmlqpegddcddepevlaavns 447
Qy      501 LH--HSLIVDTGSDSNLMLIQRLAKRVYTMVQDLSH-SGSLAFVFAIRHYSLAFDS 557
Db      448 vfelhrllagpalekylvglravrrprilyvegeanhsgtlfldfiteslhytslmids 507
Qy      558 L-----DASYGEDSPERHVVEQQLSREIRNVLAIVGSPAPT-GDVKFGSMREKI 605
Db      508 legagagsgqstacspaaaggtqgmsevylygtqicvacegaeterhelqwtstl 567
Qy      606 AOSGFRAASLAGSAAQAQSLILGMFP-SDGYTLVEBNGALKLGMKDLCLLTASAMR 660
Db      568 ggsqfapvhlgsnaykqasllalfagsgdyrveekdgciltlghwrplatsawr 623

RESULT 11
AAG22966
ID      AAG22966 standard; Protein; 240 AA.
XX
AC      AAG22966;
XX
DT      17-OCT-2000 (first entry)
XX
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 26096.
XX
KW      Protein identification; signal transduction pathway; metabolic pathway;
KM      hybridisation assay; genetic mapping; gene expression control; promoter;
XX      termination sequence.
XX      Arabidopsis thaliana.
OS      EP1033405-A2.
PN      06-SEP-2000.
XX
PD      25-FEB-2000; 2000EP-0301439.
XX
PF      25-FEB-1999; 99US-0121825.
XX      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      23-APR-1999; 99US-0130891.
PR      28-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      30-APR-1999; 99US-0132407.
PR      04-MAY-1999; 99US-0132484.
PR      05-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      06-MAY-1999; 99US-0132487.
PR      07-MAY-1999; 99US-0132863.
PR      11-MAY-1999; 99US-0134256.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      14-MAY-1999; 99US-0134370.
PR      18-MAY-1999; 99US-0134768.
PR      19-MAY-1999; 99US-0134941.
PR      20-MAY-1999; 99US-0135124.
PR      21-MAY-1999; 99US-0135353.
```

```
PR      24-MAY-1999; 99US-0135629.
PR      25-MAY-1999; 99US-0136021.
PR      27-MAY-1999; 99US-0136392.
PR      28-MAY-1999; 99US-0136782.
PR      01-JUN-1999; 99US-0137222.
PR      03-JUN-1999; 99US-0137528.
PR      04-JUN-1999; 99US-0137502.
PR      07-JUN-1999; 99US-0137724.
PR      08-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      10-JUN-1999; 99US-0138847.
PR      14-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      16-JUN-1999; 99US-0139453.
PR      17-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139450.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
PR      18-JUN-1999; 99US-0139750.
PR      18-JUN-1999; 99US-0139765.
PR      21-JUN-1999; 99US-0139817.
PR      22-JUN-1999; 99US-0139899.
PR      23-JUN-1999; 99US-0140353.
PR      23-JUN-1999; 99US-0140354.
PR      24-JUN-1999; 99US-0140695.
PR      28-JUN-1999; 99US-0140823.
PR      29-JUN-1999; 99US-0140991.
PR      30-JUN-1999; 99US-0141287.
PR      01-JUL-1999; 99US-0141842.
PR      01-JUL-1999; 99US-0142154.
PR      02-JUL-1999; 99US-0142055.
PR      06-JUL-1999; 99US-0142390.
PR      08-JUL-1999; 99US-0142803.
PR      09-JUL-1999; 99US-0142920.
PR      12-JUL-1999; 99US-0142977.
PR      13-JUL-1999; 99US-0143542.
PR      14-JUL-1999; 99US-0143624.
PR      15-JUL-1999; 99US-0144005.
PR      16-JUL-1999; 99US-0144085.
PR      16-JUL-1999; 99US-0144086.
PR      19-JUL-1999; 99US-0144325.
PR      19-JUL-1999; 99US-0144331.
PR      19-JUL-1999; 99US-0144332.
PR      19-JUL-1999; 99US-0144333.
PR      19-JUL-1999; 99US-0144334.
PR      19-JUL-1999; 99US-0144335.
PR      20-JUL-1999; 99US-0144337.
PR      20-JUL-1999; 99US-0144632.
PR      20-JUL-1999; 99US-0144884.
PR      21-JUL-1999; 99US-0144814.
PR      21-JUL-1999; 99US-0145086.
PR      21-JUL-1999; 99US-0145088.
PR      22-JUL-1999; 99US-0145085.
PR      22-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145089.
PR      22-JUL-1999; 99US-0145192.
PR      23-JUL-1999; 99US-0145145.
PR      23-JUL-1999; 99US-0145218.
PR      23-JUL-1999; 99US-0145224.
PR      26-JUL-1999; 99US-0145276.
PR      27-JUL-1999; 99US-0145913.
PR      27-JUL-1999; 99US-0145918.
PR      27-JUL-1999; 99US-0145919.
PR      28-JUL-1999; 99US-0145951.
PR      02-AUG-1999; 99US-0146386.
PR      02-AUG-1999; 99US-0146388.
```

PR	02-ANG-1999	9905-0146388
PR	03-ANG-1999	9905-0147038
PR	04-ANG-1999	9905-0147204
PR	04-ANG-1999	9905-0147332
PR	05-ANG-1999	9905-0147419
PR	05-ANG-1999	9905-0147260
PR	06-ANG-1999	9905-0147303
PR	06-ANG-1999	9905-0147416
PR	09-ANG-1999	9905-0147493
PR	09-ANG-1999	9905-0147935
PR	10-ANG-1999	9905-0148171
PR	11-ANG-1999	9905-0148319
PR	12-ANG-1999	9905-0148341
PR	13-ANG-1999	9905-0148555
PR	15-ANG-1999	9905-0149364
PR	17-ANG-1999	9905-0149375
PR	18-ANG-1999	9905-0149426
PR	20-ANG-1999	9905-0149722
PR	20-ANG-1999	9905-0149723
PR	22-ANG-1999	9905-0149902
PR	23-ANG-1999	9905-0149930
PR	25-ANG-1999	9905-0150566
PR	26-ANG-1999	9905-0150884
PR	27-ANG-1999	9905-0151065
PR	27-ANG-1999	9905-0151066
PR	27-ANG-1999	9905-0151080
PR	30-ANG-1999	9905-0151303
PR	31-ANG-1999	9905-0151438
PR	01-SEP-1999	9905-0151930
PR	07-SEP-1999	9905-0152363
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	15-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154779
PR	22-SEP-1999	9905-0155139
PR	23-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0155659
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156566
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157533
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0158926
PR	13-OCT-1999	9905-0159284
PR	14-OCT-1999	9905-0159584
PR	14-OCT-1999	9905-0159584
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160778
PR	21-OCT-1999	9905-0160760
PR	21-OCT-1999	9905-0160814
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160980
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160989
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161404
PR	26-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161459
PR	26-OCT-1999	9905-0161356
PR	26-OCT-1999	9905-0161360
PR	28-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920

PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	18.6%;	Score 651;	DB 21;	Length 240;
Best Local Similarity	56.5%;	Pred. No. 1.1e-43;		
Matches 134;	Conservative 37;	Mismatches 58;	Indels 8;	Gaps 4;

OY	427	LHNHAAHPGPRVRLTIGASMEALNKGKRLSDPADLGLPFECCAAEKAGN-VDP	4655
Dd	4	IhhlaasprPrkrlstrlrIgfssdlIastgyrrIadlsslInlpfEhnpLegllgIda	63
OY	486	EKLGVTRREAAVAHWHLNHSLVDYGDGSNTLMTLQRLAPKWVWEEDLSH--SGSP	54.3
Dd	64	sqliqrgeaxxxvnmtenrlydvngndetelrrlkpnlltveqelnpyddgsflgr	123
OY	544	FVEAIHYHSLEFDSLDASYGDESPERHNVEEQLLSREIRNVLAYGEPARTGDVKFGSME	60.3
Dd	124	fvealhyysalfdalgdglgeesgerfveeglvlt-elrnIvahgggr-----krmkwe	178
OY	604	KLAOSGFRAASLAGSAANAQSLLLGMRPPSDGYLYEENGALKLGAKDCLITTSAMR	660
Dd	179	eIstvgfivpnlgnpatqdaIlgmIpmgnygtlveengrqlrlgwklslIlttsawk	235

RESULT 12

ID AAY02539 standard; Protein; 425 AA.

AC AAY02539;

DT 16-JUL-1999 (first entry)

DE Protein encoded by the wheat Rht clone C15-1 cDNA sequence.

KW Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;
KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;;
KW paclobutrazol.

Triticum aestivum.

PN W09909174-A1

PD 25-FEB-1999.

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPT; 1999-181040/15.

XX

PT provides inhibition of the growth of plants, which inhibition is

XX

XX

CC The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC may be made by knocking out Rht or the relevant homologous gene in
CC the plant of interest. Plants may be made which are resistant to
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol

	CC	biochemical characteristics. The transcription factor may be used to
	CC	alter the structure and developmental characteristics of plants such as
	CC	soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower,
	CC	alfalfa, sugar cane, turf, banana, blackberry, blueberry, strawberry,
	CC	raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant,
	CC	grapes, honey dew, lettuce, mango, melon, onion, papaya, peas, peppers,
	CC	cucumber, spinach, squash, sweet corn, tobacco, tomato, watermelon,
	CC	tomaceous fruits and/or vegetable brassicas. Transcription factors are
	CC	key controlling elements of biological pathways and altering expression
	CC	levels of 1 or more transcription factors can change entire biological
	CC	pathways in an organism. Therefore manipulating transcription factor
	CC	levels in plants offers great potential in agricultural biotechnology
	CC	for modifying a plant's traits. Transcription factor cDNA is useful in
	CC	gene therapy.
	XX	
SQ	Sequence	587 AA:
OY	Query Match	17.0%; Score 594.5; DB 22; Length 587;
Db	Best Local Similarity	31.8%; P-adj. No. 1.1e-38;
Matches	Matches 170; Conservative	83; Mismatches 209; Indels 73; Gaps 18
OY	165 LINHREIRPCNPG-IASLEELRLSLAADRALPRPPOONALHGARAARAGL-222	
Db	81 lshactvtv-ympselyswldmll-----selnpprpj-----assngld 120	
OY	223 -TLPP-----PRLDKRRHNRPCCGCGOGEERHAPROSEKATAEETAA--268	
Db	121 vlpsrplcgfpasdydlkvlrgnalylgfraldssssnnqklkscpdmswtatc 180	
OY	269 -----AAAAAQAANAANKKEKEEBQRKRORDEEGNLITLLTLCACAEVNNADNDDA 318	
Db	181 gtcigvgigtvtltttttaaestrsvlvdsgengvrtvhaimcaaalqqnllta 240	
OY	319 HQTLTAELAEATPEGSTORVAAYFEAMASARKLVSSGLIARLPDGSRAABRLHGRVAA 378	
Db	241 ealykvqiclavagagamtkvatylafceaalrr-----lyrlappqn---qdhdlsd 289	
OY	379 AFQY-FNGISFPVKFSHTANQAIODAFEREERVNIIDLMOGLOWGFPHILASRPG 437	
Db	290 tlgmhfyetocykfkfahftangallaefegkkrrvhlidfsmgqlqwpalmqala1regg 349	
OY	438 PRPRRLTGGA----SMEALEATGRKLSDFADTLDPREFECA-VAEKAGNVDPKLGCV--490	
Db	350 prflrlglgprpadnsdhlyevoklaqlaaahveleytgtyvaansladidasmltpr 409	
OY	491 TRREAIVNMHL--HSLYDTGSDSTLTMLIORLAPKVTVWEEDLSHSGS-FLAREVEA 547	
Db	410 scteavavnsvfelhklilgrpggiekvlvgnvqikprvltfveqgsnhngvfidrfes 469	
OY	548 IHYSALFDLSIDSYGEDSPERNHVBEQQLISEIKINVLAVGSRPAT-GDVAFGSMRKELA 606	
Db	470 lhystlfdlse--gyrnsgdkvmeyvlygqclnlvaacegpddverheltsqwnrf 526	
OY	607 QSGFRASLINGSAAAOASILLMGERS-DGYTLVEENGALKGMOKDLCLTTSASAR 660	
Db	527 ssqiapahlgnsatlkgasmllyfvnsqggvyteeosngclmghwhtlrpllttsawk 581	
RESULT	15	
ID	AAG38577	standard; Protein; 517 AA.
XX		
AC	AAG38577;	
XX		
DT	18-Oct-2000	(first entry)
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 47612.	
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
XX	hybridisation assay; genetic mapping; gene expression control; promoter;	
XX	termination sequence.	

OS Arabidopsis thaliana.
XX
FN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145214.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

Db 494 dgclmlywhtrpljatsawk 513

Search completed: April 22, 2002, 19:31:46
Job time: 6031 sec

```
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 16.9%; Score 590.5; DB 21; Length 517;
Best Local Similarity 37.4%; Pred. No. 1.9e-38;
Matches 142; Conservative 71; Mismatches 140; Indels 27; Gaps 10;

```
OY 294 EBGHLHLTLLOCAFAVNADLUDAHQTLLEIAELATPFGISTORVAATYFAEAMSA RLVS 353
DB 148 engyrlvhalllaccsaavkqenltvaaalvkgiflavsqigamrkvatylaaelarrlyr 207
OY 354 SCLGLYALPFGSPAAARLHGRVAAAFQV-FNGISPFVKFSHTANQAIQEAFFERERH 412
DB 208 l-----spesqpsidhsldtltqmblyetlcpylkfahfanganalleafgkkrvh 256
OY 413 IIDLDIMQGLQPGFLHILASPGPGPPRVLTGLG----SMEALEATGKRRLSDFADTLG 468
DB 257 vidfmsqglqwpalmgalirpqpvyfrltgypapdnfdyjhvegcklahlaeah 316
OY 469 LPFERCA-VAEKAGNVDEPKGV--TREAAVAVHMLH--HSLYDVTSQSDSWTLMLIQLRLA 523
DB 317 vefeyrgftvanladdlasmlelrpselesavavsvfelhkligrpaldkvlqgvnqjlk 376
OY 524 PKVVMVQDLSH-SGSFLAFVEAIIHYYSALFDSLADSYGEDSPERHVVQQLSREIR 582
DB 377 peiftvvegeshnspltdrtfleslhytclfsle---gvpsqgdkvmseyjlgkqic 433
OY 583 NVLAVGSPART-GDYKFGSGWREKLAQSGFRASLAGSAAQASLLGMF-PSDGYTLVEE 640
DB 434 nvvaadgdpdrverheltsgvrnrtfsgagfaaahlgnafkqasmllalfnggegyrvees 493
OY 641 NGALKLGKDKCLLTASAWR 660
DB 641 NGALKLGKDKCLLTASAWR 660
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: April 22, 2002, 18:00:30 ; Search time 64.46 Seconds

(without alignments)
233.202 Million cell updates/sec

Title: US-09-265-585c-96

Perfect score: 3503

Sequence: 1 MPPPPPPPLPRYCRRCPP.....KDLCLLTASAWRPIQVPCR 668

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.5	16.9	532	4	US-09-117-853-2
2	211	6.0	282	4	US-09-117-853-8
3	196	5.6	259	4	US-09-117-853-6
4	179.5	5.1	1185	4	US-09-041-886-23
5	173.5	5.0	1958	1	US-07-945-283-2
6	171	4.9	1315	3	US-08-899-595-3
7	167.5	4.8	1248	2	US-09-080-897-2
8	167.5	4.8	1248	4	US-09-323-735-2
9	164.5	4.7	306	1	US-08-217-327-6
10	162	4.6	2205	1	US-08-093-453b-2
11	155.5	4.4	667	2	US-08-718-661-2
12	153.5	4.4	1298	2	US-08-690-473-2
13	153.5	4.4	1298	4	US-09-259-821A-2
14	153.5	4.4	1298	4	US-08-843-659-2
15	151	4.3	360	1	US-08-205-506a-2
16	151	4.3	360	5	PCT-US84-02389-2
17	149.5	4.3	635	1	US-07-832-853-2
18	148.5	4.2	286	1	US-08-382-184-3
19	148.5	4.2	286	2	US-08-641-356-3
20	148.5	4.2	286	4	US-09-132-528-4
21	148.5	4.2	286	4	US-08-875-494-3
22	148.5	4.2	325	1	US-08-382-184-2
23	148.5	4.2	325	2	US-08-641-356-2
24	148.5	4.2	325	4	US-09-132-528-2
25	148.5	4.2	325	4	US-09-132-528-3
26	148.5	4.2	325	4	US-08-875-494-2
27	148.5	4.2	332	4	US-08-818-112-53

28	147	4.2	1255	2	US-09-080-897-4	Sequence 4, Appl1
29	147	4.2	1255	3	US-08-899-595-1	Sequence 1, Appl1
30	147	4.2	1255	4	US-09-323-735-4	Sequence 1, Appl1
31	144.5	4.1	174	4	US-08-818-112-143	Sequence 143, App
32	139	4.0	908	2	US-08-476-509B-2	Sequence 2, Appl1
33	139	4.0	905	2	US-08-574-959A-9	Sequence 9, Appl1
34	139	4.0	905	4	US-09-357-014-9	Sequence 9, Appl1
35	139	4.0	1135	2	US-08-574-959A-7	Sequence 7, Appl1
36	139	4.0	1135	4	US-09-357-014-7	Sequence 7, Appl1
37	138.5	4.0	476	3	US-09-189-035-1	Sequence 1, Appl1
38	138.5	4.0	476	4	US-09-382-086-1	Sequence 1, Appl1
39	138.5	4.0	510	4	US-08-246-489-2	Sequence 2, Appl1
40	137.5	3.9	2414	5	US-08-227-536-2	Sequence 2, Appl1
41	137.5	3.9	2414	5	PCT-US95-04682-2	Sequence 2, Appl1
42	137	3.9	902	1	US-08-396-479B-6	Sequence 6, Appl1
43	137	3.9	902	1	US-08-818-823-6	Sequence 6, Appl1
44	136.5	3.9	1349	3	US-08-938-291A-6	Sequence 6, Appl1
45	136	3.9	1149	3	US-08-560-005-5	Sequence 5, Appl1

ALIGNMENTS

```
RESULT 1
US-09-117-853-2
; Sequence 2, Application US/09117853
; Patent No. 6307126
; GENERAL INFORMATION:
; APPLICANT: Harberd, Nicholas P
; APPLICANT: Peng, Jintong
; APPLICANT: Carol, Pierre
; APPLICANT: Richards, Donald E
; TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
; FILE REFERENCE: 620-45
; CURRENT APPLICATION NUMBER: US/09/117,853
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: PCT/GB97/00390
; EARLIER FILING DATE: 1997-02-12
; EARLIER APPLICATION NUMBER: GB 9602796.6
; EARLIER FILING DATE: 1996-02-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-2
```

Query Match 16.9%; Score 590.5; DB 4; Length 532;
Best Local Similarity 37.4%; Pred. No. 9.4e-40;
Matches 142; Conservative 71; Mismatches 140; Indels 27; Gaps 10;

```
QY 294 EECGHLITLLLOCAEVNNDNDHQTLLTETAEATPFGSTORVAAVPEAMSRALVS 353
      |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 163 EGVGRVLAHLLACAEVQKENTLVAELVKQIGFVAVSQIGMRVAVFAELARRIYR 222
QY 354 SCIGTVAAPLPSPAAARHGVAARAFV-ENGISDPVKESHFTANOALQFAEEERVH 412
      |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 223 L-----SPOSPTDHSLSPTLQMHYETCPYTKFAFTANQALTEFQKKRVH 271
QY 413 IIDLDIMGLQWPGLEHILASRRGPPRVRLTGIA----SNEALEATGKRISDFADTLG 468
      |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 272 VIDFSMGGQLQWPAIMQALALRPGPPVFRRLTGIRPAPDNFDYLHEVCKLAHLAEVH 331
QY 469 LPEFPCA-VAEKAQVDEPKLGV--TREDAVAVNHLH--HSIYDTGSDSTNTLQRLA 523
      |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 332 VEEYRGFAVANTLADDSMELRPSSELSVAVNSVFELHKLKIGPAIDKVLGVVQIK 391
QY 524 PKVVTWVEQDLSH-SGSFLARFEVAIHYYSALFDSLDSYGDSPERHVPGQQLSREIR 582
      |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 392 PELFVIVEQESNHNPIFLDRFTESLHYSTLFDSE---GVPSGQDKVMSEVYLGIKC 448
```

[illegible]

```

: RESULT 2
: US-09-117-853-8
: Sequence 8, Application US/09117853
: Patent No. 6307126
: GENERAL INFORMATION:
: APPLICANT: Harberd, Nicholas P
: APPLICANT: Peng, Jincrong
: APPLICANT: Carol, Pierre
: APPLICANT: Richards, Donald E
: TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
: FILE REFERENCE: 620-45
: CURRENT APPLICATION NUMBER: US/09/117,853
: CURRENT FILING DATE: 1998-08-12
: EARLIER APPLICATION NUMBER: PCT/GS97/00390
: EARLIER FILING DATE: 1997-02-12
: EARLIER APPLICATION NUMBER: GB 9602796.6
: EARLIER FILING DATE: 1996-02-12
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 282
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
: US-09-117-853-8

```

[illegible]

```

? RESULT 3
? US-09-117-853-6
? Sequence 6, Application US/09117853
? Patent No. 6307126
? GENERAL INFORMATION:
? APPLICANT: Harberd, Nicholas P
? APPLICANT: Peng, Jintong
? APPLICANT: Carol, Pierre
? APPLICANT: Richards, Donald E
? TITLE OF INVENTION: Nucleic acid encoding GAI gene of Arabidopsis thaliana
? FILE REFERENCE: 630-45
? CURRENT APPLICATION NUMBER: US/09/117,853
? CURRENT FILING DATE: 1998-08-12
? EARLIER APPLICATION NUMBER: PCT/GB97/00390
? EARLIER FILING DATE: 1997-02-12
? EARLIER APPLICATION NUMBER: GB 9602796.6
? EARLIER FILING DATE: 1996-02-12
? NUMBER OF SEQ ID NOS: 12
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 6
? LENGTH: 259

```

```

; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-117-853-6

```

Query Match	5.68;	Score 196;	DB 4;	Length 259;
Best Local Similarity	38.28;	Pred. No. 2.7e-08;		
Matches 47;	Conservative 24;	Mismatches 40;	Indels 12;	Gaps 2;

Qy	294	EGGLITLLTLLQCAAVADNDIDAHQTLTILIAELATPEFGSTORVAAYFEANASALVS	35
		: : : : : : :	
Db	146	ENGVRVLVHALLACAAVQKENVYAAELVKQIGFLVAVSQIGAMKRVATYFAEALARRTYR	205
Qy	354	SCLGIVATLPGSPAAARLHGRVAAAFQV - PNGISPFYKFSHTTANALIOEAFEREERVH	411
		: :	
Db	206	L-----SPSQSPIDHSLSDTLDMHFETCPYLKFAHFTANQALIEAFQKKRVH	255
Qy	413	IID 415	
		: :	
Db	255	IID 257	

```

1      RESULT      4
2      US-09-041-886-23
3      ; Sequence 23, Application US/09041886
4      ; Patent No. 6235872
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Bredesen, Dale E.
7      ; APPLICANT: Rahlzadeh, Sharro
8      ; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
9      ; TITLE OF INVENTION: Polypeptides and Methods of Use
10     ; NUMBER OF SEQUENCES: 72
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Campbell & Flores LLP
13     ; STREET: 4370 La Jolla Village Drive, Suite 700
14     ; CITY: San Diego
15     ; STATE: California
16     ; COUNTRY: United States
17     ; ZIP: 92122
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Floppy disk
20     ; COMPUTER: IBM PC compatible
21     ; OPERATING SYSTEM: PC-DOS/MS-DOS
22     ; SOFTWARE: PatentIn Release #1.0, Version #1.25
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/09/041,886
25     ; FILING DATE:
26     ; CLASSIFICATION:
27     ; ATTORNEY/AGENT INFORMATION:
28     ; NAME: Campbell, Cathryn A.
29     ; REGISTRATION NUMBER: 31,815
30     ; REFERENCE/DOCKET NUMBER: P-LJ 2626
31     ; TELECOMMUNICATION INFORMATION:
32     ; TELEPHONE: (619) 535-9001
33     ; TELEFAX: (619) 535-8949
34     ; INFORMATION FOR SEQ ID NO: 23:
35     ; SEQUENCE CHARACTERISTICS:
36     ; LENGTH: 1185 amino acids
37     ; TYPE: amino acid
38     ; TOPOLOGY: linear
39     ; MOLECULE TYPE: protein
40     ; US-09-041-886-23

```

```

Query Match Similarity 5.1%; Score 179.5; DB 4; Length 1185;
Best Local Similarity 21.5%; Pred No. 4.8e-06;
Matches 166; Conservative 69; Mismatches 259; Indels 279; Gaps 37

QY 2 PPPPP-----PPLPRGRCRPP-----H 21
    |||||      |||
Db 442 PPPPPGGLLANSNAPRRPSTGAOSTANRRSTNNNNHOOOOOOOOOONHGN 501
    |||||
QY 22 LPPPPSSNNHFL-----LHFLHQLDHOEAAAAAMVKRRASOMDLRPPRRHVT----- 70
    |||||

```



```

Db 502 GPPPPGAPPPLEGGSSHHAN--PYAMSSLSGLRPPGPAHLPPHPSQVSYSOAGPN 559
QY 71 -----GDSVYTAANAAG-----VGGSGAP-----SSASQALPAPPTOLHQL 107
Db 560 GPPVSSSSSSSSSSQSCSPCHSPSPQOGAPPPPPYPTVYTSATLSTYATVASS 619
QY 108 PPAFOHHAPEVDYVPAHPAANAAGGATASTTAVYDGIIRDTIGSSGAASVITQLIH 167
Db 620 PAGYKTASP-----PGPPYCKRAPSPPAYKTATPPGY-----KPGSPSFRGTGTPP 666
QY 168 NVREIHPNPGSLASLELRSLAADPAPLPPPPPOQHALLHGRAPAPGL-TLPP 226
Db 667 GYKGTSPAGGTFK-----PGSPYVGPGLPP-----AGPSGLSLPP 705
QY 227 PPLPDRKRHHPPPCQOOOEEPP--HPAPSPKAPT-----261
Db 706 PPAAP-----ASGPPLSATQIKQEPAEYETPPESPVPAPASPPPKVVDYPSNASQSAF 761
QY 262 -----AETAAAAAQAQAAAAAKERKEEORRQRDEE 295
Db 762 NKHLDRGFNSCARSDLYFVPLEGSKLAKKRADLYEKVRRAEQAREREREREREKE 821
QY 296 -----GLHLLTLLOCAEA-VNADNLDA-HQTLLEFIATPFP--GTSTQVAAIYFA 344
Db 822 RERERELERSVKLAQEGRAPECPSLGVPYHR-----PPFPGSAVATVPPYLG 872
QY 345 -EAMASRLVSSCLGYA--PLPPGSPAARLH-----GRVAAAFQVFNGISPEVFESH 394
Db 873 PDTALMTLSF-----YARPHMSPGN--RNHPFYPLGAVDPGLLGYN--VPALXSSD 922
QY 395 FTANAOIOEAERERER-----VHIIDIMQIOWPGLFHI-----LASRPGGPPRY 441
Db 923 PAAREEREAERDLRLKPGFEVKPELEPLHGVPGPLDPPFRHGLALOPGP--979
QY 442 RLTLGASMEALATGKRLSDFADTLGLPFEFCVAEAKGNV-----DDEKLGVPREEA 495
Db 980 -----GLHPPFPSPSLG-PLERERLALAGALRPMDSIVAEKLAERKQHA 1023
QY 496 VAVHMLHSLDYVTSNTLMILQRLAPKYVTWVEODLSHSGSFL-ARVEAIIHYSA- 553
Db 1024 ERVAGL-----GNP-----LARIOMLVNPHNHOSHINSHLHLHOODAIHAASAS 1070
QY 554 ---LFDLI-DASYGEDSP-----ERHVEQQLISREIRVLA 586
Db 1071 VHPILIDPLASGSHLTRIPYAGTLPLNPLRPHLEHNEVLRHOLFAPARYDLPA 1123

```

```

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4011 ext. 513
TELEFAX: 309-685-4128
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1958 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-945-283-2

Query Match
Best Local Similarity 21.4%; Pred. No. 3e-05; Length 1958;
Matches 100; Conservative 26; Mismatches 124; Indels 217; Gaps 21;

QY 2 PPPPP-----PPLPYCRRCRPPHLP--PPSSPHFLYHLQ-----40
Db 321 PGPPTSTSHNNHNGRPTSP--RPSTSHQDPPGGGPPSAETH--NNHNDPPGGP 375
QY 41 -----LDHQBAAAAAVKRRASDMDLPPRRRNVGTGLSDVTAATAAGVGGSGAPSSAS 94
Db 376 PSTSHNNHNDP-----PGGPPSPPPR-----PSTSS 403
QY 95 ADLRLPQLHOLPRAFOHHAPEVDYVPAHPA-----AHAQAGGATASTTAVYDGIIRD 150
Db 404 -----SSHQGPSTRRPPRQRRPPRPPSPKISTRTAGSENTYQTL-----447
QY 151 IIGSSGAASVITQLIHNVREIT-HPCNPGSLASLELRSLIAA-----DPAPLPPPQ 204
Db 448 -----FSHSNKLFHSHPGEGEG--DRGTAGGEGDDPPRPPSPR 488
QY 205 PQOHALLHAPRAARGLTLPPRPLDKRKHENHPPCQOOOEEHPHAPOSP-----257
Db 489 P-----PPLRPPPPPPPPQ-----PPAGSARRRRRGGGGPPRGRR 530
QY 258 ---KAPTAEATAAAAAA-----AAAAAAKERKEE 286
Db 531 RGGKRRRAESTERAAAADAEEEDGDEDEDAEDBGRDEGEGRGAGCGESESSE 590
QY 287 QRRQ--RDEE-----GLHLLTLLOCAEVNADNLDDAHQTLLEIAE 327
Db 591 SSRAEGAPRSAGQVGVAGVLGLVROGLHDPERAGRAPVAADADDLHR-----643
QY 328 LATPFGTSTQRYAATFAEASRLVSSCLGYAPLPPGSPAARLHG 374
Db 644 -----GRVLPVLAG-----PPGARGFVGLHG 664

```

```

RESULT 5
US-07-945-283-2
; Sequence 2, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911

```

```

RESULT 6
US-08-899-595-3
; Sequence 3, Application US/08899595
; Patent No. 6111072
; GENERAL INFORMATION:
; APPLICANT: Narumiya, Shun
; APPLICANT: Takahashi, No. 6111072uak1
; TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE
; TITLE OF INVENTION: ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,595
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-242701
; FILING DATE: 26-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 9-90170
; FILING DATE: 25-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Stephen A. Bent
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 049441/0112
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1315 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-595-3

```

```

Query Match 4.9%; Score 171; DB 3; Length 1315;
Best Local Similarity 20.0%; Pred. No. 2,7e-05;
Matches 117; Conservative 53; Mismatches 176; Indels 238; Gaps 25;

```

```

QY 2 PPPPPPLTPYCRRCPPPLP-----PPSSPNHFLHYLHQLDHEAAAAAMVKR 55
DB 656 PPPPPPLPGVCISPPSLGGTAISPPPLS----- 689
QY 56 PASDMDLPPRRHYTGLSDVTAAGVGGSGAPSSASQLP--ALPTQLHQLPPAFQ 112
DB 690 --GDATIPP-----PLPE-----GVGIPSPSS--LPGTAIP-----PPPL 724
QY 113 HHAPVDVPAHRAAHQAAGEATSTAWDGIIRDIITSSGGAANSITOLIHNVREI 172
DB 725 PGSARIPPPPLP-----GSAG----- 742
QY 173 IHPCHGLASLLELRSLAADPARLPPOOHALLHGAPAAAPAGLTLPPLPD 232
DB 743 IPPPPPLPG-----EAGMPPPP-----LPGGPIPPPPPPPG 777
QY 233 KRHEHPPCCQQOQOQEEP--HPAPOSKPAP-----TAETA- 266
DB 778 GGIPIPPPGMGMPPPPGFVPAAPVLPFGLTTPKLYKPEVOLRRPNMSKLVADLSQ 837
QY 267 -----AAAAAQAATAAAK-----EKKEBRKRRQDEBGLHL 299
DB 838 DCFWTKVKEDRENNELFAKLTLTFSAQTKTSKAKKDEGGEKKSVQKKVKE----- 891
QY 300 LTLLOCAEAVNADNL-----DDAHOITLLEIAELATPFGTSTORAAVFAEAMS 348
DB 892 ----LKVLDSKTAQNLSTFLSGFRMPYOEIKNVILLEVNE-AVLTESMIONLIKMPPEBQ 946
QY 349 ARLVSSCGLYAPLPGPS-----AARLHGVAAP-----QVENGISPFVKFSHF 395
DB 947 LKMSLELDEYDILAESQFGVMGTVPRLRPLNALILFKLQFSEQYEN-----IKPEIV 1001
QY 396 TANQAIOGAFAERERVAHIIDIMOGLOWPGLFHILASRPGGPPRRVRLTIGASMELEA 455
DB 1002 SVTAACELRKSESEFSNLETTLL-----VGNYNMAGSRNAGAFNISFL-CXLRDTKS 1055
QY 456 TGRRL-----SDFADTLGLPFEFCVAVAEKAGVDEPKL 488
DB 1056 TDQKKTLLHFLAELCENDYPDYLFKPPDELAHV-EKASRVSANL 1098

```

```

RESULT 7
US-09-080-897-2
; Sequence 2, Application US/09080897
; Patent No. 598574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Weissh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UM97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-080-897-2

```

```

Query Match 4.8%; Score 167.5; DB 2; Length 1248;
Best Local Similarity 19.6%; Pred. No. 4.9e-05;
Matches 114; Conservative 53; Mismatches 171; Indels 243; Gaps 25;

```

```

QY 2 PPPPPPLTPYCRRCPPPLP-----PPSSPNHFLHYLHQLDHEAAAAAMVKR 55
DB 600 PPPPPPL-----PPPLPGTAISPPPLS----- 625
QY 56 PASDMDLPPRRHYTGLSDVTAAGVGGSGAPSSASQLP--ALPTQLHQLPPAFQ 112
DB 626 --GDATIPP-----PLPE-----GVGIPSPSS--LPGTAIP-----PPPL 660
QY 113 HHAPVDVPAHRAAHQAAGEATSTAWDGIIRDIITSSGGAANSITOLIHNVREI 172
DB 661 PGSARIPPPPLP-----GSAG----- 678
QY 173 IHPCHGLASLLELRSLAADPARLPPOOHALLHGAPAAAPAGLTLPPLPD 232
DB 679 IPPPPPLPG-----EAGMPPPP-----LPGGPIPPPPPPPG 713
QY 233 KRHEHPPCCQQOQOQEEP--HPAPOSKPAP-----TAETA- 267
DB 714 GGIPIPPPGMGMPPPPGFVPAAPVLPFGLTTPKLYKPEVOLRRPNMSKLVADLSQ 773
QY 268 -----AAAAAQAATAAAK-----EKKEBRKRRQDEBGLHL 302
DB 774 DCFWTKVKEDRENNELFAKLTLTFSAQTKTKKDEGGEKKSVQKKVKE----- 824

```



```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1298
; TYPE: PRT
; ORGANISM: HERPES VIRUS, TYPE 1
US-09-259-821A-2

```

```

Query Match      4.4%; Score 153.5; DB 4; Length 1298;
Best Local Similarity 20.4%; Pred. No. 0.0007;
Matches 140; Conservative 69; Mismatches 209; Indels 267; Gaps 30;

```

```

QY 53 KRKPAS--DMDLPPRRHVTGDLSDVTAANAAGVSGAPSSASQALPALTQHLQLPRA 110
DB 6 KORGSGPTDGPPTPPSPDRD-----ERGALGMCATETEGGDDP-----45
QY 111 FOHNAPEVDVPHAPAPAAHQAQGEATSTTAWVDGIIRDIISSGGAASITQLIHNV 170
DB 46 --DHDPD-----HPHDDARRDGRAPAACTD-----AGEDAGDASVPRQLAL---86
QY 171 EIIHPCNPGLASLLELRSLAADPAFLP-----PPQ 204
DB 87 -----LASVVEEAVRTIPRPDPAASPRTPAPFRADDDGDGDEYDADAAGDAPAR 137
QY 205 -POOHALLHGAPAAAGFLPPLPPPLDOKRREHPPPCQQQQQEEPHAPQSPKAPTAE 263
DB 138 GREHEAPLRGAYPDPTDRLSPRPAPQPRRRRH-----GRWRPSAS 178
QY 264 ETAAAAAQAQAAAAAAKEREKREORRKORDEGLHLTLTLLOCAEAVNADNLDAHQTL 323
DB 179 STSDSGSSSSSSSSSSSSSDE---DEDDDG-----NDAADHARE---216
QY 324 EIAELATPPTGISTQRYAAYFAEAMSAARLVSSCLGLYAPLPSPG-----AAARL- 372
DB 217 -----ARAVGRGPSSAAPAPGR-----TPPPGPPPLSEAPKPRAAARTP 258
QY 373 -----HGRVAAAFQVFNGISPFVKFSHTANQ-----AIOEAFEREERYHI 413
DB 259 AASAGRIERRRARAAGVAGDATG-----RFTAGQPRRVELDADATSGAFYARYRDGY 310
QY 414 IDDDIMOGLOWPGLFHILASRPGPPRVRLTGLGASMEAL-----EATGKRL 460
DB 311 V-----SGEPWPG-----AGPPPGRVLYGLGDSRPGMGAPAEAEARRFEASGAPA 359
QY 461 SDFADTLG-LPEFCAVAEKAQNVDPKLGVTREAVAVHMLH-----SLYD 507
DB 360 AWAAPELGDAAQOYALITRLTYPPDAEMG-----WLONPVVPEVDVALDQACFR 409
QY 508 VTGSDSNTLMLIORLAPKVVTWEQDLSSHGSFLA--REVEAIIHYSSA-----553
DB 410 ISGAARNS-----SSFTIGSVARAAPVHLGYAAMAAGRFMGGLAHAAAVAMSRRYDRAQ 462
QY 554 ---LFDSLDASYGDSPEKRVNVEQLLSREIRNVL-AVGSPARTGD-----YKFGSWRE 603
DB 463 KGFLLTSLIRRYA-----PILLARENAALTGAAGSPGAGADEGVAVAANAAPGE 511
QY 604 KLAQSGFRAASLAGS-----AAQAS 624
DB 512 RAAPAGYGAGLIALGRLSAPAS 536

```

```

RESULT 14
US-08-843-659-2
; Sequence 2, Application US/08843659
; Patent No. 6218103
; GENERAL INFORMATION:
; APPLICANT: Leopaldi, Rosario
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS US3 AND ICPI AS
; TITLE OF INVENTION: INHIBITORS OF APOPTOSIS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee

```

```

; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Ver. 2.1
; CURRENT APPLICATION DATA:
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: AKB-519
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1298 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-843-659-2

```

```

Query Match      4.4%; Score 153.5; DB 4; Length 1298;
Best Local Similarity 20.4%; Pred. No. 0.0007;
Matches 140; Conservative 69; Mismatches 209; Indels 267; Gaps 30;

```

```

QY 53 KRKPAS--DMDLPPRRHVTGDLSDVTAANAAGVSGAPSSASQALPALTQHLQLPRA 110
DB 6 KORGSGPTDGPPTPPSPDRD-----ERGALGMCATETEGGDDP-----45
QY 111 FOHNAPEVDVPHAPAPAAHQAQGEATSTTAWVDGIIRDIISSGGAASITQLIHNV 170
DB 46 --DHDPD-----HPHDDARRDGRAPAACTD-----AGEDAGDASVPRQLAL---86
QY 171 EIIHPCNPGLASLLELRSLAADPAFLP-----PPQ 204
DB 87 -----LASVVEEAVRTIPRPDPAASPRTPAPFRADDDGDGDEYDADAAGDAPAR 137
QY 205 -POOHALLHGAPAAAGFLPPLPPPLDOKRREHPPPCQQQQQEEPHAPQSPKAPTAE 263
DB 138 GREHEAPLRGAYPDPTDRLSPRPAPQPRRRRH-----GRWRPSAS 178
QY 264 ETAAAAAQAQAAAAAAKEREKREORRKORDEGLHLTLTLLOCAEAVNADNLDAHQTL 323
DB 179 STSDSGSSSSSSSSSSSSSDE---DEDDDG-----NDAADHARE---216
QY 324 EIAELATPPTGISTQRYAAYFAEAMSAARLVSSCLGLYAPLPSPG-----AAARL- 372
DB 217 -----ARAVGRGPSSAAPAPGR-----TPPPGPPPLSEAPKPRAAARTP 258
QY 373 -----HGRVAAAFQVFNGISPFVKFSHTANQ-----AIOEAFEREERYHI 413
DB 259 AASAGRIERRRARAAGVAGDATG-----RFTAGQPRRVELDADATSGAFYARYRDGY 310
QY 414 IDDDIMOGLOWPGLFHILASRPGPPRVRLTGLGASMEAL-----EATGKRL 460
DB 311 V-----SGEPWPG-----AGPPPGRVLYGLGDSRPGMGAPAEAEARRFEASGAPA 359
QY 461 SDFADTLG-LPEFCAVAEKAQNVDPKLGVTREAVAVHMLH-----SLYD 507
DB 360 AWAAPELGDAAQOYALITRLTYPPDAEMG-----WLONPVVPEVDVALDQACFR 409
QY 508 VTGSDSNTLMLIORLAPKVVTWEQDLSSHGSFLA--REVEAIIHYSSA-----553
DB 410 ISGAARNS-----SSFTIGSVARAAPVHLGYAAMAAGRFMGGLAHAAAVAMSRRYDRAQ 462

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 19:31:50 ; Search time 72.24 Seconds

(without alignments)
704.383 Million cell updates/sec

Title: US-09-265-585c-96

Sequence: 1 MPPPPPPPLPPIYCRPCPP.....KDLCLITPASAMKPIQVPCRR 668

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR-68:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1865	53.2	653	2	T47581 SCARECROW1 - Arabi
2	1861	53.1	653	2	T51244 SCARECROW protein
3	601	17.2	511	2	G96688 hypothetical prote
4	593.5	16.9	587	2	D84426 hypothetical prote
5	590.5	16.9	533	2	H86282 protein F10B6.34 I
6	572	16.3	523	2	T51475 RGA-like protein -
7	570.5	16.3	526	2	E96542 scarecrow-like pro
8	561	16.0	658	2	D96656 hypothetical prote
9	548	15.6	413	2	G84462 probable SCARECROW
10	544.5	15.5	593	2	E86347 hypothetical prote
11	536.5	15.3	482	2	E96540 hypothetical prote
12	499.5	14.3	306	2	T51236 scarecrow-like pro
13	498	14.2	352	2	T51234 scarecrow-like 7 (
14	470.5	13.4	542	2	T46142 scarecrow-like 7 (
15	457	13.0	483	2	T10552 hypothetical prote
16	442	12.5	718	2	T45597 probable SCARECROW
17	439.5	12.5	583	2	T45597 scarecrow-like pro
18	423.5	11.9	375	2	C71441 probable SCARECROW
19	417.5	11.2	1336	2	T02736 scarecrow-like pro
20	399.5	11.4	808	2	T51241 scarecrow-like pro
21	398	11.4	808	2	T51232 scarecrow-like pro
22	368.5	10.5	531	2	T04722 hypothetical prote
23	367	10.5	640	2	B84887 probable SCARECROW
24	366.5	10.5	325	2	T51235 scarecrow-like pro
25	340	9.7	733	2	T51239 scarecrow-like pro
26	340	9.7	623	2	T47874 scarecrow-like pro
27	317	9.0	486	2	E85433 SCARECROW-like pro
28	314	9.0	313	2	T51242 Sc11 protein (limp
29	309	8.8	410	2	T45848 hypothetical prote

30	299.5	8.5	558	2	T01343 hypothetical prote
31	291	8.3	378	2	T51237 scarecrow-like pro
32	287	8.3	1585	2	T31611 hypothetical prote
33	226.5	6.5	261	2	T51243 Sc11 protein (limp
34	208.5	6.0	1357	2	T29265 hypothetical prote
35	204.5	5.8	590	1	T35297 probable dihydrol
36	201	5.7	1048	2	T31425 C-terminal domain-
37	195.5	5.6	1201	2	G86441 unknown protein (l
38	193.5	5.5	1151	2	T18535 high molecular mas
39	192.5	5.5	1541	2	T16755 hypothetical prote
40	192.5	5.5	1541	2	T02831 AAA protein L4171.
41	192	5.5	1892	2	T18314 hypothetical prote
42	189	5.4	1184	2	S50832 atrophin-1 - human
43	187	5.3	3421	1	W2BEB6 367k tegument prot
44	184.5	5.3	907	2	E96636 hypothetical prote
45	183.5	5.2	707	2	A46302 PTB-associated spl

ALIGNMENTS

RESULT 1	T47581	SCARECROW1 - Arabidopsis thaliana
N:Alternate names:	protein F24B22.180	
C:Species:	Arabidopsis thaliana (mouse-ear cress)	
C:date:	20-Apr-2000	#sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession:	T47581	
R:Blocker:	H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quettier, F.; Salanoubat, M.	
A:Reference number:	223016	
A:Accession:	T47581	
A:Status:	preliminary	
A:Molecule type:	DNA	
A:Residues:	1-653 <BLO>	
A:Cross-references:	EMBL:AL132957	
C:Genetic:		
A:Map position:	3	
A:Introns:	511/2	
A>Note:	F24B22.180	
Query Match	53.2%	Score 1865; DB 2; Length 653;
Best Local Similarity	57.4%	Pred. No. 4e-102;
Matches	405; Conservative	66; Mismatches 122; Indels 112; Gaps 18;
QY 4 PPPPPPL-----TPYCRCPPLPPPPSSPHFLHYHQDHQAAAVRRRP 56		
DB 12 PPPHSLRTTSSGSSNNKRGPP-PPPP-----LVNVRRL 48		
QY 57 ASDM-----DLPPRR-H-VTGDLSDVT-----AAAAAGVGSQAPG--SASQ 96		
DB 49 ASMSNPNPYNNSSRRPRVSHLDSNVTVTPQQPSTLAATVYSSQNPPLSVCGPSG 108		
QY 97 LPLALPQLHQLPPARQHAPEYDVPAHPAPAAHAAGCATSTTAWDGIITRTIGSSG 156		
DB 109 LPVFPp-----DRGRNVMSVQPMW-----ODSSSSASPTVWVDATIRDIHSS- 154		
QY 157 GAAVSTTOLIHVRETIHPCNPGIASILELRSLAADPA---PLP----- 200		
DB 155 -TSVSLPQLIOWRDITIFPCNPLGALLEYRLRSMLDLPSSSDPSPTQTEPPLTQISNN 213		
QY 201 --PPPOQHALLHGAPAAAPAGLTLPPPLPDRRHENHPCCQOOOEHPAPQSPK 258		
DB 214 PSPPOQOOQH-----QOOQHKKPPPPPTQOQERE-----NSSTD 248		
QY 259 APAAETAAAAA-AAQAAAAAAKERKEBQRKQDEGSLHTLLTLLCCAEVADNIDDD 317		
DB 249 AAPPEPTVATVPAAVQNTAAELRERKEIKRQKDEGSLHTLLTLLCCAEVADNLEE 308		
QY 318 AHQTLLEIALELTPGTSTQRYAAFAFAMSRILVSSGLGYAPLPSPSPAARHG-RV 376		

```

Db 309 ANKLLLEISQISTPYGTSAORVAAYFSEAMSARLLNSCLGIYAALP--SRMMPQTHSLKM 366
OY 377 AAAFQVFNGISPFVKESHFTANQAIQEAFFEREERHIIIDLMQGLQMPGLFHILLASRPG 436
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 VSAFQVFNGISPLVKFSHTANQAIQEAFFEKEDSVHIIIDLMQGLQMPGLFHILLASRPG 426
OY 437 GPPRVRLTGLGASMEALATGKRLSDPDTGLPFEEFCVAEKAQNDPEKLGVTREAV 496
   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 427 GPPHVRLLGTSMELQATGKRLSDFADKGLPFEEFCPLAEKVGNDTERLNVRRKREAV 486
OY 497 AVHMLHSLYDVTGSDNTMLIQRLAPKYVTWVEDLSHSGSFELARFEVAIHYYSALFD 556
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 AVHMLQHSLYDVTGSDAHTLMLQRLAPKYVTWVEDLSHAGSFLGRFEVAIHYYSALFD 546
OY 557 SLDSAYGEDESPERRHVEEQQLSREIRNVLAAGPARTGDKVKSREKLQSGFRASLSA 616
   || ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 547 SLGASYGESEERHVEEQQLSKEIRNVLAAGPSRSGEVKFSREKMOQCGFKGISLA 606
OY 617 GSAAQASLLGMFPSPDGYTLVEENGALKLGKDKLCLLTASAMRP 661
   :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 GNAATQATLLGMFPSPDGYTLVDNGTLKLGKDKLSLTASAMTP 651

```

RESULT 2

```

TS1244
SCARACROW protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 28-Jul-2000
C:Accession: TS1244
R:di Laurencio, L.; Wyszocka-Diller, J.; Malamy, J.E.; Pysk, L.; Helariutta, Y.; Freshour
submitted to the EMBL Data Library, July 1996
A:Reference number: 225338
A:Accession: TS1244
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1 653 <DIL>
A:Cross-references: EMBL:U62798; PIDN:AA06318.1
C:Genetics:
A:Gene: SCARACROW1
A:Introns: 511/2

```

```

Query Match 53.1%, Score 1861; DB 2; Length 653;
Best Local Similarity 57.3%, Pred. No. 6.9e-102;
Matches 404; Conservative 66; Mismatches 123; Indels 112; Gaps 18;

OY 4 PPPPPPL-----TPYCRRCRPHLP PPPSSPNHFLHLHQHLDQEAAMAVRRKR 56
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 PPHSPPLRTTSSGSSSSNNRCPPP--PPPP-----LVWVRKRL 48
OY 57 ASDM-----DLPPRR--H-VTGDLSDT-----AAAAGVGSAGS---SASAQ 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49 ASEMSNSNDYNNSSRPKRVSHILDSNNYVTPQOPSLTAAYVSSPNPPLSVCGFSG 108
OY 97 LPALFTQHLQLEPAFQHAPEVDVPAHPAPAAHAQAGEATATTTAVDGIIRDISSSG 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 LPVFPS-----DRGRVNMMSVQPM-----ODSSSSASPTVWDAILRILHSS- 154
OY 157 GAAVSTQQLINHVREIHPNGLASILELRSLADPA-----PLP----- 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 155 -TSVSIPLQIQNVRIIFPCNPNIGALLEYRSLMLDPSSSSDPSQTFEPPLYQISNN 213
OY 201 --PPQPOQHALLHGAPAAAGLTLPPPLPDKRHHENPPCQOQOQOEPRHAPSPK 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 PSPPQOQOQOHO-----QOQOQOKPPPPPIQOQERE-----NSSTD 248
OY 259 APTAETAAAA-AAQAAAAAAAKERKEQRRKORDEEGHLTLTLQCAEAVNADLND 317
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 APPQETATVAVQVOTATALREKKEIRKQKODEGLHLLTLQCAEAVADNLEE 308
OY 318 AHQTLLEIAELATPGCTSTORVAAAFAMASARLVSSGLGAPLPSPSPAAALHG- RV 376
   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 309 ANKLLLEISQISTPYGTSAORVAAYFSEAMSARLLNSCLGIYAALP--SRMMPQTHSLKM 366

```

```

OY 377 AAAFQVFNGISPFVKESHFTANQAIQEAFFEREERHIIIDLMQGLQMPGLFHILLASRPG 436
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 367 VSAFQVFNGISPLVKFSHTANQAIQEAFFEKEDSVHIIIDLMQGLQMPGLFHILLASRPG 426
OY 437 GPPRVRLTGLGASMEALATGKRLSDPDTGLPFEEFCVAEKAQNDPEKLGVTREAV 496
   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 427 GPPHVRLLGTSMELQATGKRLSDFADKGLPFEEFCPLAEKVGNDTERLNVRRKREAV 486
OY 497 AVHMLHSLYDVTGSDNTMLIQRLAPKYVTWVEDLSHSGSFELARFEVAIHYYSALFD 556
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 AVHMLQHSLYDVTGSDAHTLMLQRLAPKYVTWVEDLSHAGSFLGRFEVAIHYYSALFD 546
OY 557 SLDSAYGEDESPERRHVEEQQLSREIRNVLAAGPARTGDKVKSREKLQSGFRASLSA 616
   || ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 547 SLGASYGESEERHVEEQQLSKEIRNVLAAGPSRSGEVKFSREKMOQCGFKGISLA 606
OY 617 GSAAQASLLGMFPSPDGYTLVEENGALKLGKDKLCLLTASAMRP 661
   :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 GNAATQATLLGMFPSPDGYTLVDNGTLKLGKDKLSLTASAMTP 651

```

RESULT 3

```

G96688
hypothetical protein T27F4.10 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96688
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chn, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G96688
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <STO>
A:Cross-references: GB:AE005173; NID:q10092507; PIDN:AG12907.1; GSPDB:GN00141
C:Genetics:
A:Gene: T27F4.10
A:Map position: 1

```

```

Query Match 17.2%, Score 601; DB 2; Length 511;
Best Local Similarity 34.6%, Pred. No. 4.2e-28;
Matches 146; Conservative 70; Mismatches 150; Indels 56; Gaps 10;

OY 273 QAAAAAAKEREKQRRKOR-----DESGHLLTLTLQCAEAVNAD 313
   :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 RAIPGSAYVRPDHVTNRSKRTRIESELSTRSVVLDSEQGVRLVAALLCAEAVOON 166
OY 314 NLDAHQTLLEIAELATPGCTSTORVAAAFAMASARLVSSGLGAPLPSPSPAAARLH 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 167 NLKLADALVNHVGLLASSQAGAMRKAYTFABELARIT-----RTY 208
OY 374 GRVAAAFQVFNGL-----SPFKESHFTANQAIQEAFFEREERHIIIDLMQGLQMP 425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 PRDVALSFSDFLQIHFPESCYLEKFAHFTANQAIQEAFFAEKAVHVIDLGLNGLQMP 268
OY 426 GLFHILLASRPGGPPRVRLTGLGASMEALATGKRLSDPDTGLPFEEFCVAEKAQND 484
   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 269 ALIQALALRNGPPRPLRGIGSLTDIQEVGKKLQGLASTIGVNPFEFSIKLNNLSDK 328
OY 485 PEKLGV--TREAAVAVHMLH--HSLYDVTGSDNTMLIQRLAPKYVTWVEDLSHSGS-F 540
   ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 329 PEMLDTRPGLESVAVSVFELHRLAHPSIDKFLSTISIRPDIMTVVEQDANNGITYF 388

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 19:34:35 ; Search time 71.83 Seconds

(without alignments)
340.973 Million cell updates/sec

Title: US-09-265-585c-96

Sequence: 1 MPPPPPPPLTPYCRRCRPP.....KDLCLLTASAMRPVPCRC 668

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	217	6.2	2715	1 TRX2_HUMAN	O9um66 homo sapien
2	201	5.7	1048	1 SRA4_RAT	O63627 rattus norv
3	199.5	5.7	290	1 TRX2_MOUSE	O08550 mus musculu
4	198	5.7	1157	1 SRA4_HUMAN	O95104 homo sapien
5	193	5.5	428	1 FXB2_MOUSE	O64733 mus musculu
6	187	5.3	1183	1 DRPL_RAT	P54258 rattus norv
7	187	5.3	3421	1 TEGU_HAVER	P28955 equine herp
8	183.5	5.2	707	1 SEPO_HUMAN	P23246 homo sapien
9	179.5	5.1	1185	1 DRPL_HUMAN	P54259 homo sapien
10	178	5.1	721	1 YK82_MYCTU	P10690 mycobacteri
11	171	4.9	620	1 EXTN_TBOAC	P13983 nicotiana t
12	169.5	4.8	1733	1 VNDA_PRYKA	P33485 pseudorabie
13	168.5	4.8	633	1 LA17_YEAST	O12446 saccharomyc
14	167.5	4.8	1248	1 DIAL_HUMAN	O60610 homo sapien
15	167.5	4.8	1790	1 SEPA_HUMAN	P78621 emeritella
16	167.5	4.8	2142	1 BAT2_HUMAN	P48634 homo sapien
17	165.5	4.7	1083	1 T2D3_HUMAN	O00268 homo sapien
18	164.5	4.7	306	1 EXTN_DAVCA	P06599 daucus caro
19	164.5	4.7	564	1 ZYX_MOUSE	O62523 mus musculu
20	164.5	4.7	901	1 Y298_HUMAN	O15016 homo sapien
21	163.5	4.7	572	1 ZYX_HUMAN	O15942 homo sapien
22	162.5	4.6	497	1 WAS2_HUMAN	O9y665 homo sapien
23	162	4.6	426	1 EXLP_TBOAC	O03211 nicotiana t
24	162	4.6	2205	1 POLN_RUBVT	P13889 rubella vir
25	161	4.6	502	1 WAS3_HUMAN	O9uy66 homo sapien
26	161	4.6	982	1 CBLE_HUMAN	O13191 homo sapien
27	161	4.6	1132	1 BAT3_HUMAN	P46379 homo sapien
28	160.5	4.6	339	1 CSP_PLABE	P06915 plasmodium
29	160.5	4.6	1603	1 PSC_DROME	P35820 drosophila
30	160	4.6	347	1 CSP_PLABE	P23093 plasmodium
31	160	4.6	1794	1 YAVI_SCHPO	O10172 schizosacch
32	159.5	4.6	505	1 WAS1_HUMAN	O00401 homo sapien
33	158.5	4.5	559	1 WAS1_HUMAN	O92558 homo sapien

ALIGNMENTS

```

RESULT 1
ID      TRX2_HUMAN          STANDARD; PRT; 2715 AA.
AC      O9UM66; O9UK25; O95836; O9Y669; O9Y668; O15022;
DT      20-AUG-2001 (Rel. 40, Created)
DT      20-AUG-2001 (Rel. 40, Last sequence update)
DE      TRITHORAX HOMOLOG 2 (MIXED LINEAGE LEUKEMIA GENE HOMOLOG 2 PROTEIN).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_TaxID=9606;

RN      [1]
RP      SEQUENCE FROM N.A. (LONG ISOFORM).
RA      Angrand P.O., Valvatne H., Jeanmougin F., Adamson A.,
RA      van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,
RA      Lamerdin J., Chambon P., Losson R., Stewart A., Asland R.;
RT      "Mammalian trithorax- and ASH1-like proteins: putative chromatin
RT      regulators which contain PHD fingers and SET domains.";
RL      Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.

RN      [2]
RP      SEQUENCE FROM N.A. (LONG ISOFORM).
RA      Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,
RA      Garcia E., Kyle A., Ramirez M., Stilwagen S., Ganes J., Danganan L.,
RA      Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
RA      Carrano A.V.;
RT      "Sequence analysis of a 1 Mb region in human 19q13.1";
RL      Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.

RN      [3]
RP      SEQUENCE OF 816-2715 FROM N.A. (LONG ISOFORM).
RC      TISSUE=Brain;
RA      MEDLINE=97349984; PubMed=9205841;
RA      Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA      Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT      "Prediction of the coding sequences of unidentified human genes. VII.
RT      The complete sequences of 100 new cDNA clones from brain which can
RT      code for large proteins in vitro.";
RL      DNA Res. 4:141-150(1997).

RN      [4]
RP      SEQUENCE OF 111-2715 FROM N.A. (LONG ISOFORM).
RC      TISSUE=Testis; and Leukocyte;
RA      MEDLINE=20105772; PubMed=10637508;
RA      Hundsman D.G., Chin S.-F., Mulleris M., Batley S.J., Collins V.P.,
RA      Wiedemann L.M., Aparicio S., Caldas C.;
RT      "MLL2, the second human homolog of the Drosophila trithorax gene, maps
RT      to 19q13.1 and is amplified in solid tumor cell lines.";
RL      Oncogene 18:7975-7984(1999).

RN      [5]
RP      PARTIAL SEQUENCE FROM N.A. (LONG AND TRUNCATED ISOFORMS).
RC      TISSUE=Placenta, and Bone marrow;
RA      MEDLINE=99339983; PubMed=10409430;
RA      Fitzgerald K.T., Diaz M.O.;
RT      "MLL2: A new mammalian member of the trx/MLL family of genes.";
RL      Genomics 59:187-192(1999).

-i- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.

```

CC	-1 SUBCELLULAR LOCATION: NUCLEA (BY SIMILARITY).
CC	-1 ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A TRUNCATED FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC	-1 TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHEST LEVELS IN TESTIS.
CC	ALSO FOUND IN BRAIN, BONE MARROW, HEART, MUSCLE, KIDNEY, PANCREAS,
CC	SPLEEN, THYMUS, PROSTATE, OVARY, INTESTINE, COLON, PERIPHERAL
CC	BLOOD LYMPHOCYTES, AND PLACENTA.
CC	-1 DISEASE: OFTEN AMPLIFIED IN PANCREATIC CARCINOMAS.
CC	-1 SIMILARITY: BELONGS TO THE MYTHORAX FAMILY OF TRANSCRIPTION FACTORS.
CC	-1 SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC	-1 SIMILARITY: CONTAINS 1 SET DOMAIN.
CC	-1 SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch)
CC	or send an email to license@isb-sib.ch)
CC	-----
DR	EMBL; AJ007041; CAB45385.1; -
DR	EMBL; AD000671; - NOT_ANNOTATED_CDS.
DR	EMBL; AB002302; BAA20763.2; -
DR	EMBL; AF186605; AAD56420.1; -
DR	EMBL; AF104918; AAD17932.1; -
DR	EMBL; AF105279; AAD26113.1; -
DR	EMBL; AF105280; AAD26112.1; -
DR	InterPro; IPR000637; AT-hook.
DR	InterPro; IPR000389; FYRICH_C.
DR	InterPro; IPR003888; FYRICH_N.
DR	InterPro; IPR001965; PHD.
DR	InterPro; IPR002965; P-rich_extensions.
DR	InterPro; IPR003616; PostSET.
DR	InterPro; IPR002857; ZnF-CXXC.
DR	InterPro; IPR001841; ZnF-fing.
DR	Pfam; PF00628; PHD; 3.
DR	Pfam; PF00856; SET; 1.
DR	Pfam; PF02008; zf-CXXC; 1.
DR	SMART; SM00384; AT_hook; 1.
DR	SMART; SM00542; FYRC; 1.
DR	SMART; SM00541; FYRN; 1.
DR	SMART; SM00249; PHD; 4.
DR	SMART; SM00508; PostSET; 1.
DR	SMART; SM00184; RING; 1.
DR	SMART; SM00317; SET; 1.
DR	PROSITE: PS50280; SET; 1.
KM	DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;
KW	Transcription regulation; Alternative splicing.
FT	DNA_BIND 37 44
FT	DNA_BIND 110 117
FT	DNA_BIND 357 365
FT	ZN_FING 1203 1252
FT	ZN_FING 1253 1303
FT	ZN_FING 1337 1396
FT	DNA_BIND 1449 1471
FT	DNA_BIND 2586 2715
FT	DNA_BIND 26 37
FT	DNA_BIND 248 255
FT	DNA_BIND 362 398
FT	DNA_BIND 402 771
FT	DNA_BIND 808 812
FT	DNA_BIND 1963 1970
FT	DNA_BIND 2251 2259
FT	VARSPLIT 532 582
FT	-----
PT	VARSPLIT 583 2715
FT	CONFLICT 834 834
FT	CONFLICT 941 941
FT	-----
FT	MISSING (IN ISOFORM TRUNCATED).
FT	R -> E (IN REF. 5).
FT	S -> Y (IN REF. 5).
FT	-----
FT	VSARRSVIKTPREKDEDPKPPEVSEVPYLRRPITSP
FT	VSQEPAPVS -> PLSOSLLPMTLIOSTSGCWAAETTS
FT	ACDSPMLSPLLRRCPRLTGQL (IN ISOFORM
FT	TRUNCATED).

FT	CONFLICT	1317	1317	E -> Q (IN REF. 5).
FT	CONFLICT	1362	1362	H -> Y (IN REF. 5).
FT	CONFLICT	1438	1438	D -> N (IN REF. 5).
FT	CONFLICT	2622	2622	D -> H (IN REF. 5).
SO	SEQUENCE	2715	AA: 293511 MW: C0615B981BBEB7BF	CRC64:

Query Match	6.2% ; Score 217 ; DB 1 ; Length 2715 ;
Best Local Similarity	23.5% ; Pred. No. 0.00032 ;
Matches 129 ;	Conservative 35 ; Mismatches 187 ; Indels 198 ; Gaps 27 ;

QY	1	MPRP-----PPRLPYCRRCRPPRLPPRPSSPNHLLHYLDHDEAAAAANVR	53
DB <td>401</td> <td>LPRLPLPPAPSPRPRLPPSTSPRLCPPPP-----</td> <td>434</td>	401	LPRLPLPPAPSPRPRLPPSTSPRLCPPPP-----	434
QY <td>54</td> <td>KRPASDML-----PPRRKRVTDGLSPVTAANAAGVCGSAPSSASQALPALTOLHOLPRA</td> <td>110</td>	54	KRPASDML-----PPRRKRVTDGLSPVTAANAAGVCGSAPSSASQALPALTOLHOLPRA	110
DB <td>435</td> <td>--VPSPRLPPSPRPRAOEEGEESPVPVPAATCSKKRGP-----PLTSQARERPEA</td> <td>485</td>	435	--VPSPRLPPSPRPRAOEEGEESPVPVPAATCSKKRGP-----PLTSQARERPEA	485
QY <td>111</td> <td>FOHNAPEVDVNAHNPAAHNAOAGEATASTAWWDGILRIDIGSSGAAVSTOLDHNVR</td> <td>170</td>	111	FOHNAPEVDVNAHNPAAHNAOAGEATASTAWWDGILRIDIGSSGAAVSTOLDHNVR	170
DB <td>486</td> <td>--RAPRBTSP--PIPTPSTATGCRPDESP-----VAKSTFTFLNIR</td> <td>525</td>	486	--RAPRBTSP--PIPTPSTATGCRPDESP-----VAKSTFTFLNIR	525
QY <td>171</td> <td>ELIHPENGLASLLELRSLAADPAPLP-----PPPOQHALLGAP</td> <td>215</td>	171	ELIHPENGLASLLELRSLAADPAPLP-----PPPOQHALLGAP	215
DB <td>526</td> <td>QFIMFVVSARSRSRYIKTPRRFMEDEP--PKRPKEVSPVLRPIITTSPPVQE-----PAP</td> <td>579</td>	526	QFIMFVVSARSRSRYIKTPRRFMEDEP--PKRPKEVSPVLRPIITTSPPVQE-----PAP	579
QY <td>216</td> <td>AAAPRGLLP--PPPLDCKR-----HEPPRCOOQOEERHAPAPSPKA</td> <td>259</td>	216	AAAPRGLLP--PPPLDCKR-----HEPPRCOOQOEERHAPAPSPKA	259
DB <td>560</td> <td>VPSPRAATPTPTVPRLPEKRRSILRETFKWTSLTRLP-----PPAPPPRAPSPRA</td> <td>636</td>	560	VPSPRAATPTPTVPRLPEKRRSILRETFKWTSLTRLP-----PPAPPPRAPSPRA	636
QY <td>260</td> <td>P-TAEETAAAAAQAQAAAA-----AKERKEORRKORDEEGHLTL</td> <td>303</td>	260	P-TAEETAAAAAQAQAAAA-----AKERKEORRKORDEEGHLTL	303
DB <td>637</td> <td>PATSSRRPLLRAPQFTSEAHKLYESVLTPLPLGAEAEPEPEPPADDSF-----</td> <td>688</td>	637	PATSSRRPLLRAPQFTSEAHKLYESVLTPLPLGAEAEPEPEPPADDSF-----	688
QY <td>304</td> <td>LOCAR-----ATNADLDHANOHTLEIABLAT-----PFGT-----STORVAAEFA</td> <td>344</td>	304	LOCAR-----ATNADLDHANOHTLEIABLAT-----PFGT-----STORVAAEFA	344
DB <td>669</td> <td>---APRERAPGRN-----HLSLPFAFVYTTPYKAEVSPHGAPARLSNGPOTQAOLOPL</td> <td>741</td>	669	---APRERAPGRN-----HLSLPFAFVYTTPYKAEVSPHGAPARLSNGPOTQAOLOPL	741
QY <td>345</td> <td>EAMSAIIVSSCGLYAP--LRPGSP-----AAARLHGVAAPAOVFNGISPPYKFSHT</td> <td>396</td>	345	EAMSAIIVSSCGLYAP--LRPGSP-----AAARLHGVAAPAOVFNGISPPYKFSHT	396
DB <td>742</td> <td>QALQQLPQALPPRPOQLPPSPQOMPLEKARIAG-----VGSLLP-----</td> <td>785</td>	742	QALQQLPQALPPRPOQLPPSPQOMPLEKARIAG-----VGSLLP-----	785
QY <td>397</td> <td>ANQALQ-----AFERERHIIIDLDLMQGLQMPGLPHILASRGGPRRRLTGLASMEAL</td> <td>453</td>	397	ANQALQ-----AFERERHIIIDLDLMQGLQMPGLPHILASRGGPRRRLTGLASMEAL	453
DB <td>786</td> <td>--SGVEEMKFSILKRAKQVLRIDQOQOQK-----VAASPLSP-----GGOMEV</td> <td>829</td>	786	--SGVEEMKFSILKRAKQVLRIDQOQOQK-----VAASPLSP-----GGOMEV	829
QY <td>454</td> <td>EATGRISD</td> <td>462</td>	454	EATGRISD	462
DB <td>830</td> <td>AGAVKQISD</td> <td>838</td>	830	AGAVKQISD	838

RESULT 2		
SR44_RAT	STANDARD ;	PRT ; 1048 AA.
AC	063627 ;	
DT	30-MAY-2000 (Rel. 39, Created)	
DT	30-MAY-2000 (Rel. 39, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	CTD-BINDING SR-LIKE PROTEIN R44 (FRAGMENT).	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Hippocampus;	
RX	MEDLINE=96293459; PubMed=8692929;	
RA	Yuvraj A., Patturajan M., Litingtung Y., Joshi R. V., Gentile C.,	
RA	Gehara M., Corden J. L.	
RT	"The C-terminal domain of the largest subunit of RNA polymerase II	
RT	interacts with a novel set of serine/arginine-rich proteins."	
RL	Proc. Natl. Acad. Sci. U.S.A. 93:65975-65980(1996).	

CC -1- FUNCTION: MAY ACT TO PHYSICALLY AND FUNCTIONALLY LINK
 CC TRANSCRIPTION AND PRE-MRNA PROCESSING (BY SIMILARITY).
 CC -1- SUBUNIT: INTERACT WITH THE REPETITIVE C-TERMINAL DOMAIN (CTD) OF
 CC RNA POLYMERASE II (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U49058; AAC52660.1; -
 CC InterPro: IPR000504; RRM.
 CC Pfam: PF00076; rrm; 1.
 CC SMART: SM00360; RRM; 1.
 CC PROSITE: PS0102; RRM; 1.
 CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 CC DR RNA-binding.
 CC KW NON-TER 1 1
 CC FT DOMAIN 3 6 POLY-PRO.
 CC FT DOMAIN 155 158 POLY-PRO.
 CC FT DOMAIN 422 496 RNA-BINDING (RRM).
 CC FT DOMAIN 630 638 POLY-PRO.
 CC FT DOMAIN 655 661 POLY-PRO.
 CC FT DOMAIN 834 857 POLY-GLN.
 CC FT DOMAIN 834 857 POLY-GLN.
 CC SEQUENCE 1048 AA; 113701 MW; FSCF8BD309D683E CRC64;

Query Match 5.7%; Score 201; DB 1; Length 1048;
 Best Local Similarity 23.0%; Pred. No. 0.00081;
 Matches 80; Conservative 31; Mismatches 99; Indels 138; Gaps 16;

QY 1 MPPPPPPPLTPYC-----RCPPPHLP--PPPPSSPNHFLHYLQLDHQEAAAMV 52
 DB 629 VPPPPPPPP--PFWGPGFNPMLPPLPPPPPPPPPPPPPPPPPPPPPPPPPPPP 664
 QY 53 RKRASDMDLPPR-----RHYTGDLSDVTAAGVSGAPSSASQALPALPTQLH 105
 DB 665 ---EVS---IPPHHTPPISTIPNLVSGARGNAESADSAKMGSGAP-----PAAPTSL- 710
 QY 106 QLPRAFGHNAEDVPAHPARAHAQAGGEXTASTAWDGIIRDIIGSSGGAIVSTQL 165
 DB 711 PTPPVYTG-----FVSLGTQGVAPGP-----VIGLQAPSTGLLG----- 745
 QY 166 IHNREIHPNCGLASLELRSLAADPAALP-----PPQQOHALLHGAPA 216
 DB 746 -----RPLGLPLQR-----PCGMPPHLPORFPMMPRPMPPHMMHRRPP 784
 QY 217 AAPAGLTPRP-----PP-----LPD-----KRHEHPPPCOOQ 246
 DB 785 PGGGFAPRRHGKGRPPRGPRVGRGGMGLGPGRGSGEDRGRQOQPOQOQOQO 844
 QY 247 QDEHPAPQSPKAPTAETATAAAAAAQAQAAKKEKREKRRQORDE 294
 DB 845 QO 892

RESULT 3
 TRX2_MOUSE STANDARD; PRT; 290 AA.
 ID TRX2_MOUSE
 AC O08550;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRITHORAX HOMOLOG 2 (MW DOMAIN BINDING PROTEIN 7) (FRAGMENT).
 GN TRX2 OR WBP7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97315177; PubMed=9171351;
 RA Bedford M.T., Chan D.C., Leder P.;
 RT "Fbp MW domains and the Abi SH3 domain bind to a specific class of
 RT proline-rich ligands";
 RL EMBO J. 16:2376-2383(1997).
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
 CC FACTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U92455; AAC53192.1; -
 CC MGD: MGI:109565; Wbp7.
 CC InterPro: IPR000637; AT_hook.
 CC Pfam: PF02178; AT_hook; 1.
 CC SMART: SM00384; AT_hook; 1.
 CC DNA-binding; Nuclear protein; Transcription regulation.
 CC KW NON-TER 1 1
 CC FT DNA_BIND 1 7 A.T HOOK (BY SIMILARITY).
 CC FT DOMAIN 4 35 ASP/GLU-RICH (ACIDIC).
 CC FT DOMAIN 45 77 POLY-PRO.
 CC FT DOMAIN 253 263 POLY-PRO.
 CC FT NON-TER 290 290
 CC SEQUENCE 290 AA; 31202 MW; C49B5565FPC8D2DF CRC64;

Query Match 5.7%; Score 199.5; DB 1; Length 290;
 Best Local Similarity 25.4%; Pred. No. 0.00023;
 Matches 75; Conservative 22; Mismatches 91; Indels 107; Gaps 12;

QY 2 PPPPPPPPLTPYCRRCPPPHLP--PPPPSSPNHFLHYLQLDHQEAAAMVKKRPSDMD 61
 DB 46 PVPSPPPPL-----PPSTSPPPASP----- 67
 QY 62 LPPRRHVTGDLSDVTAAGVSGAPSSASQALPALPTQLHLP-----AFQHH 115
 DB 68 LPPP-----VSPPLSPPPYPA--PEKQESSPLVVPATCSRRKGR 106
 QY 116 PEVDVPAHPARAHAQAGGEXTASTAWDGIIRDIIGSSGGAIVSTQLHNHREIHP 175
 DB 107 PPL-TPSQARERAAASGEGTISPSPNPTTSGSLEDSPVIVKSTFLKNIFQFIMP 165
 QY 176 CNEGLASLELRSLAADPAALP-----PPQQOHALLHGAPAAAPAG 221
 DB 166 VVASRSSRVYIKTPRRRMDPDP--PKPKVEASIVRPVATSPAPDE-----FVVPSSPPR 219
 QY 222 LTLPP--PPPLDPKRR-----HEHPPCQOQOQOQDEHPAPQSPKAP 260
 DB 220 VPPPSTPVLPPEKRSILREPTFTWTSILRELPP-----PPAPPPAPSPPPAP 269

RESULT 4
 SRA4_HUMAN STANDARD; PRT; 1157 AA.
 ID SRA4_HUMAN
 AC Q95104; Q9UPM1; Q9ULP8;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CTD-BINDING SR-LIKE PROTEIN R44 (FRAGMENT).
 GN KIAA1172.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NX NCBI_TaxID=9606;

FT	DNAI1	750	756	POLY-PRO.
FT	DNAI1	967	976	POLY-GUN.
FT	VAMP3L1C	776	798	STGAINEDTTKDLSTGNPIPTV -> L (IN ISOFORM 2).
FT	CONFICT	270	274	MISSING (IN REF. 3).
FT	CONFICT	293	293	T -> I (IN REF. 1).
FT	CONFICT	296	296	T -> A (IN REF. 1).
FT	CONFICT	309	309	T -> A (IN REF. 1).
FT	CONFICT	312	312	V -> L (IN REF. 4).
FT	CONFICT	562	562	H -> Y (IN REF. 1).
FT	CONFICT	570	570	L -> V (IN REF. 1).
FT	CONFICT	581	581	N -> T (IN REF. 2).
FT	CONFICT	610	610	GVT -> CVI (IN REF. 1).
FT	CONFICT	614	616	MDT -> CDN (IN REF. 1).
FT	CONFICT	622	622	L -> M (IN REF. 1).
FT	CONFICT	626	626	C -> Y (IN REF. 1).
FT	CONFICT	818	836	AVPRAAPTNLTPTPVTOCV -> TCCTHESAHPSNCPACLP C (IN REF. 1).
FT	CONFICT	837	866	SLLTGQVAARVPVIGLOEPTSTGLGARPGI -> VSILGKL FFFLIISFCCPRPPRTTHDQVRFE (IN REF. 3).
FT	CONFICT	854	854	A -> E (IN REF. 1).
FT	CONFICT	1149	1149	S -> G (IN REF. 1).
SO	SEQUENCE	1157 AA:	126846 MM;	176A/78998BFB01B CRC64;

Query Match Best Local Similarity 5.7%; Score 198; DB 1; Length 1157;
Matches 86; Conservative 24; Mismatches 98; Indels 172; Gaps 17;

OY	1	MRRP-----PPRPPLTRXCR-----RCRPHLR--PPRSSRNHFLNLHOLDHOEA	46
DB	716	IPRFEGFGVPRPPRRPFRLRGFNHMLRGEFLDRGGRRPTTP-----	759
OY	47	AAAMVKKRKRPASDMDLRPP-----RRNTGDLS-----DYTAAMAAGV	84
DB	760	-----PVS-----IPRHTRPISINSTIAGINEDTKDISIGNIPRTIVSGARGMAE	807
OY	85	GGSGARPSSASQALRALPQLHOLRRAPFNHNAREVDVNRANHRARAAMOAGGEGATASTAMV	144
DB	808	SGDSVKMKGSVAPRAPRPNL--PTRVTV-----PVS-----	837
OY	145	DGIIRIDIGSSGGAAVSTITQLIHNVREIINPCNSGLASLELYKSLAADPA--PL---	199
DB	838	-----LLGTGVAV-----PGRVIGLARSTGLGARPGILRPLQR	872
OY	200	----PP-----PROPQHALLHGABAARAGILTLP-----	227
DB	873	PGMRPRLHORPLRMKRPMRNMMNGRRPBGAFMRRGMKGRRPHGFRVARGMP	932
OY	228	-----PRLRDKNRHHNPSSOOOEEBRNAPOSKAKATTEGTAATAAAAAA	279
DB	933	GLGGGRRGRRGERDRDGKQDFRQDPQQO---PQPARQORQQOQPPSQDRPPPTQQ	988
OY	280	AKERKEQRRKO---RDEE	295
DB	989	QPQGFRRNDNRQOFNSGRDQE	1008

RESULT 5
FXB2_MOUSE AC 064733;
ID FXB2_MOUSE STANDARD; PRT; 428 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
GN FORKHREAD BOX PROTEIN B2 (TRANSCRIPTION FACTOR FKX-4).
GN FOXB2 OR FKX4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
PP SEQUENCE FROM N.A.

Matches 118; Conservative 30; Mismatches 154; Indels 300; Gaps 24;

```

OY 2 PPPPP-----PLPPYCRRC---PPPHLPP-----25
    |||||
Db 165 PPSPPPPSIPQPSGPEPHPSVPTGYNAMDEPPTSRLLQGGPPGAPPHPOLYPSGA 224
OY 26 -----PPSSNHLHYLHQLDHOEAAAAMVKKRPASDMDLPPRRHVTGDLSDVTA 78
    |||||
Db 225 GGGVLSPGMPG-----KGGAAASSVCPRSSGCKPHPP-----TTP 260
OY 79 AAAAGVGGSGAPSSASQDLPALPTQLHQLPP--AFQHHAPVDV-----121
    |||||
Db 261 IPISSSGASGAPRPNTPVGAGNLPSPAPPTPHTPMLPPPALPLNNASASBP 320
OY 122 -----AHPAPAHQAQGG-----EAT 137
    |||||
Db 321 MCAQPIPGHLPSPHAMGGMSGLPPEPKGPTLAPSRPLPPASSABGPPMRPYSSCS 380
OY 138 ASTTAVDGIIRDIISSSGAAVSITQLIHNVREIHPNCNGLASLELRSLADPA 197
    |||||
Db 381 SSSVA-----ASSSSAATS-----QYRASQTLPSY-----PH 408
OY 198 PLPPPPQ-----POQHALLHGAPAAAPAGLLPPPPP-----LPDKRHH--236
    |||||
Db 409 SFPPTSMVSNOQPKYTPQSLPSQAVWSQGP-----PPPPYGRLLPPNNTHPG 459
OY 237 -----EHP-PC-----QOQOQEEHPAPQ-----SPKAPTAETAAAAAQA 274
    |||||
Db 460 PEPPTGCGSTAHRRAPAHNNHQOQOQRPQRPQOHHNGNSGRRPFGA-----507
OY 275 AAAAAAKERKEQRKKQDEBGLHLTLTLLOCAEAVNANDIDAHQTLLEIATLPFGT 334
    |||||
Db 508 -----YHPRLSSNSHNH-----PYN-524
OY 335 STORVAAYFAEAMARSCLGLYAPLPPGSPAARLHGVAAPQVENGISPFYFESH 394
    |||||
Db 525 -----MSPSLGSLRPYPPGPAHLPPSHGQVSTSQAGPMG--PVSSSS 565
OY 395 FTANQAIQEAFAERERHIIIDIMQGLQWGLFPHILASRGPPRYRLTGLGASMEALE 454
    |||||
Db 566 NSSGSSSQAAVSCS---HPSSSGPQGASYP-----PPVPVPIITSSATLTSTVI 611
OY 455 AT 456
    ||
Db 612 AT 613

```

RESULT 7
TEGU_HSVB STANDARD; PRT; 3421 AA.

AC P28935;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE LARGE TEGUMENT PROTEIN.
GN 24.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirinae.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1."
RL Virology 189:304-316(1992).
CC -1- FUNCTION: TEGUMENT PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHV-1 24, EBV BFLF1, HSV-1 64, VZV 22, AND HCMV UL48.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M86664; AAB02459.1; -
DR PIR: G36797; WZBBB6.
DR HSSP: P04002; 1WFA.
SQ SEQUENCE 3421 AA; 367078 MW; 5075EEF4739BB7AC CRC64;

Query Match 5.3%; Score 187; DB 1; Length 3421;
Best Local Similarity 26.7%; Pred. No. 0.017;
Matches 82; Conservative 32; Mismatches 119; Indels 74; Gaps 13;

```

OY 1 MPPPPPPITPYCCRRCPPHLP PPPSSPNHFLHYLHQLDHOEAAAAMVKKRPASDM 60
    |||||
Db 2570 LPAPPLPOSTSKAASGATQSDGKTLTLDVFKTQSKDKVP-----LPQTSKASG-----2607
OY 61 DLPPRRHVTGDLSDVTAAGVGGSGAPSSASQDLPALPTQLHQLPPAPQHNAPEVDY 120
    |||||
Db 2608 --PPPTLPPAPPLPOSTSKAASGATQSDGKTLTLDVFKTQSKDKVP-----VPPTDK 2659
    |||||
OY 121 PAHPAPAHQAQGEATASTTAVDGIIRDIISS-----SGGAVSITQLIHNVREIHH 174
    |||||
Db 2660 PSTTTPAALKQS--DASKPPTAIIQH--QQKIGTPVTPKDSQDKPTDNASAPVGSVPTP 2715
OY 175 PCNPGLASLELRSLADPAFLPPPP-----QOQHALL--LHGA-----PAAAPAG 221
    |||||
Db 2716 DGTGG-----AKPPKDAIPVDDTKQPVRRSLQVKGGRPYRISLGPFK 2761
OY 222 LTLPPPLPDKRHRHPPPP-----CQOQOQEEHPAPQSKAPTAETTA-AAAAAQA 277
    |||||
Db 2762 FTGPPGTYTP---VHGLPPSDSNVTQSTKEPKPAVETPAAPASAAAAPAAAPAKSAAA 2818
OY 278 AAKERK 284
    ||
Db 2819 PAAAPAK 2825

```

RESULT 8
SFPQ_HUMAN STANDARD; PRT; 707 AA.

AC P23246; P30808;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SPLICING FACTOR (PSF) (DNA-BINDING P52/P100 COMPLEX, 100 KDA SUBUNIT).
DE BINDING FACTOR, PROLINE- AND GLUTAMINE-RICH (POLYRIMIDINE TRACT-
DE FACTOR) (PSF) (DNA-BINDING P52/P100 COMPLEX, 100 KDA SUBUNIT).
GN SFPQ OR PSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RP TISSUE=Fetal brain;
RC MEDLINE=93194059; PubMed=8449401;
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor."
RT Genes Dev. 7:393-406(1993).
RN [2]
RP SEQUENCE OF 312-707 FROM N.A.
RP TISSUE=Fetal skeletal muscle;
RC MEDLINE=90091812; PubMed=2480877;
RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RT "Cloning and characterization of a myoblast cell surface antigen
RT defined by 24.1b5 monoclonal antibody."
RL Development 105:723-731(1989).
RN [3]
RP SEQUENCE OF 48-68 AND 213-246.

```

RX MEDLINE-93176127; PubMed-8439294;
RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
RT Purification and characterization of a DNA-binding heterodimer of 52
RT and 100 kDa from HeLa cells.;
RL Biochem. J. 290:267-272(1993).
CC -1- FUNCTION: ESSENTIAL PRE-MRNA SPLICING FACTOR REQUIRED EARLY IN
CC SPLICOSOME FORMATION. BINDS TO THE MAMMALIAN POLY(PYRIMIDINE
CC TRACTS). FORMS A COMPLEX WITH THE POLY(PYRIMIDINE TRACT-BINDING
CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
CC -1- SUBUNIT: HETERO-DIMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RRM).
CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
CC SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
CC ECTOKINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70944; CAAS0283.1; -
DR EMBL: X16850; CAA34747.1; -
DR PIR: A43557; A43557.
DR PIR: S29770; S29770.
DR HSSP: P19339; ISXL.
DR MIM: 605199; -
DR InterPro: IPR00504; RRM.
DR Pfam: PF00076; rrm; 2.
DR SMART: SM00360; RRM; 2.
DR PROSITE: PS0102; RRM; 2.
DR PROSITE: PS00030; RRM_RNP.1; 1.
KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
KW Alternative splicing.
FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
FT REPEAT 9 27 3 x 3 AA REPEATS OF R-G-G.
FT REPEAT 19 21 1.
FT REPEAT 25 27 2.
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VARSPLIC 663 707 KTERRGQGGAGVGCGGPRGPGTPTAGYGRGREYEG
FT CONFLICT 243 243 PKKKPRF -> VRMIDVG (IN SHORT ISOFORM).
FT FT G -> R (IN REF. 3).
SQ SEQUENCE 707 AA; 76149 MW; 6D8D5EA95E235847 CRC64;

```

Query Match 5.2%; Score 183.5; DB 1; Length 707;
 Best Local Similarity 21.8%; Pred. No. 0.0045;
 Matches 143; Conservative 43; Mismatches 188; Indels 281; Gaps 32;

```

QY 1 MPPPP-----PPPLTPCRRCRPPHLPSPSSPNHFLHLHQLDHOEAAAMVAK 54
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 60 ITPPPHHQQQQQQPPRPOOPRQQ-PPPHOPPPH----- 91
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 55 RPASDMDLPPRRHHVTGDISDVTAAAGVG-----SGA-PSSASNDLP 98
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 92 QPHQQQQPPPPDSSSKPVVAGGPGAPGVGSAPASSAPATPTTSGAPGSGPGPR 151
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

```

QY 99 ALPTOLHOLPPAFONHAEVDVPAHAPAAHQAQGEATTAWDGIIRDITGSSGA 158
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 152 TPBPVATSApG-----APPPTPSSGVPTTPQAGP----- 184
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 159 AVSTIQLHNHVEITHPCNPGLASLLELRSLAADPAPLPPPPQ-----POOHALL 212
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 185 -----PPP-----AAVPGPGPGKPGGPGKPGKPG 214
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 213 GAPAAAPAGLTPPPPLPPDKR-----RHEPPPGQQQQGEEHPAPSPAPATAE 264
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 215 GKPPGGGGLSTPGGHPKPPPHNGGEPGPGKOHHPYHQHHQGPFGPGG----- 266
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 265 TAAAAAAAQAAAAAAKKEKEORRQRDEGLHL-LTLLQCEAA-----VN 311
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 267 -----RSEE-----KISDSGEFKANLSLRPRPKTYTCRCFLVGNLP 305
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 312 ANNDHAHOTLLELALATP-----TGTSTQRAAYFAEAMSRVLSCLGLYAP 361
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 306 ADITDEFKRL--FAKYGEPEGVFINKGKGFILKESRA-LAEIAKAEI----- 352
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 362 LPPGSPAAR-LHGRV-----AAAFVFNGLSPVKKFSHTANQIDAEFER---EEVHI 413
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 353 --DTPMKGRQLRYFPATHAALSVRN-LSPYV-----SNELLEAFSGPPIERAVI 403
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 414 IDLDIMQGLWGLFPHILLASRPGGPPRVRLTGLG-----ASMEALEATGKRISD---FADT 466
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 404 VD-----DRGRSTGKGIIVEFASAPPAARKAPKERSGEGFLTT 440
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 467 LGLPFEPCAVAKAGNVD-----PEKLVGTRREAVAVVHLHSLDYTGSSNTLMILQIR 521
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 441 TPRP---VIVEPRLDQLDEDEGLPEKL-----ACKNPMYOKER 474
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 522 LAPKVTVWEQDLSHSGFLAFVEAIIHYVSLPDSIASYGEDSPERHVEQOL 576
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 475 ETP-----PRFAQHGT-----EYEQSRKMSLDEM---EKOQNEQVEKKN 512
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

```

RESULT 9
DRPL_HUMAN STANDARD; PRT; 1185 AA.
ID DRPL_HUMAN AC P54259;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ATROPHIN-1 (DENTATORUBRAL-PALLIDOLUYSIAN ATROPHY PROTEIN).
GN DRPLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Brain;
RX MEDLINE-95144175; PubMed-7842016;
RA Nagatuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K.,
RA Inoue T., Yamada M.;
RT "Structure and expression of the gene responsible for the triplet
RT repeat disorder, dentatorubral and pallidoluyisan atrophy (DRPLA).";
RL Nat. Genet. 8:177-182(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-96262314; PubMed-8965642;
RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RA Kidwai A.S., Ashworth R.G., Ross C.A.;
RT "DRPLA gene (atrophin-1) sequence and mRNA expression in human
RT brain";
RL Brain Res. Mol. Brain Res. 36:219-226(1996).
RN [3]
RP SEQUENCE OF 470-725 FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE-93315145; PubMed-8325628;
RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;

```


"Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms.";
RL Genomics 16:572-579(1993).
CC -1- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC Ovary, Testis and Prostate. Lower levels are detected in the
CC Liver, Thymus and Leukocytes.
CC -1- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC DISEASE.
CC -1- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
CC PALLIDOLYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC NUCLEUS, RUBRUM, GLOBUS PALLIDUS AND LYSYS/BODY. CLINICAL FEATURES
CC ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC IN THE FOURTH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D31840; BAA06626.1; -;
DR EMBL: U23851; AAB50276.1; -;
DR EMBL: L10377; -; NOT_ANNOTATED_CDS.
DR HSSP: P00651; ILRA.
DR MIM: 125370; -;
DR InterPro: IPR002951; Atrophin.
DR PRINTS: PR01222; ATROPHIN.
KW Triplet repeat expansion; Polymorphism.
FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).
FT 302 305 POLY-PRO.
FT 376 382 POLY-SER.
FT 386 397 POLY-SER.
FT 442 447 POLY-PRO.
FT 479 483 POLY-HIS.
FT 484 497 POLY-GLN.
FT 504 507 POLY-PRO.
FT 564 574 POLY-SER.
FT 704 707 POLY-PRO.
FT 802 815 ARG/ALA-RICH (MIXED CHARGE).
FT 816 827 ARG/GLU-RICH (MIXED CHARGE).
FT 925 934 ARG/GLU-RICH (MIXED CHARGE).
FT 934 94 MISSING (IN REF. 2).
FT 94 94 Y -> H (IN REF. 2).
FT 333 333 Y -> I (IN REF. 2).
FT 339 339 M -> I (IN REF. 2).
FT 541 541 P -> T (IN REF. 3).
FT 1028 1028 G -> A (IN REF. 2).
FT CONFLICT 1028 1028
SQ SEQUENCE 1185 AA; 124785 MW; 56C306267331C005 CRC64;

Query Match 5.1%; Score 179.5; DB 1; Length 1185;
Best Local Similarity 21.5%; Pred. No. 0.013;
Matches 166; Conservative 69; Mismatches 259; Indels 279; Gaps 37;

QY 2 PPPPP-----PLTPRCRCRCPP-----H 21
DB 442 PPPPPGRLANSNANRPPRPSTGAQSTANRPVSTNNHHNQOQOQOQOQOQNHNG 501
QY 22 LPPPPSSPNHFL-----LHYHOLDHOEAAAAMVRRKRPASDMLPRPRRHVT----- 70
DB 502 SGPPPGGFRPHLEGGSSHNAN--PYAMSPFSLGSLRPRREGRAHLPRHNSOVSTISQAGPN 559
QY 71 -----GDLSDVYAAAAAG-----VGSAGD-----SSAAQPLPALTQHL 107
DB 560 GPPVSSSSSSSSSTSGSIPCSHPSPSGPGAGAPYPPRPVYVTTTSATLSTVATVASS 619
QY 108 PPAFHNAPEVDVPAHRAAHAAQAGBATASTTAWDGIITDIIGSSGAANSITQLIH 167

DB 620 PAGYKTASP-----PGPPYGRAPSPGAKTATPPCY-----KQSPSPSFTCTPP 666
QY 168 NWREIHPNCPGLASLELRSLSLAADPAPLPPPPQOQHLLGAPAAAPAGL-TLPP 226
DB 667 GYKGTSPGAPGTFK-----PGSPVCGPLPP-----AGBSGLSP 705
QY 227 PPLPLPKRRHNEPPPCQOQOQEEP--HPAPOSPAPRT----- 261
DB 706 PPAAP-----ASGPPLSATIQKQPAEYETPSVPYRASPSPPKVVDVPSHASQARF 761
QY 262 -----AETAAAAAAQAAAAAAKERKEEORRORDEE 295
DB 762 NKHLDFNSCARSDLYFVPLEGSKLAKKRADLVKVRREARQAREREKEREREKE 821
QY 296 -----GHLTLTLQCEA-VNADNLDA-HQTLLELAELATPF--GTSQRYAAYFA 344
DB 822 RERERELEERSVKYLAQEGRAVECPSLGVPVHR-----PPRESAVAIVPPLG 872
QY 345 -EAMSAFLVSSCLGYA--PLPPGSPAARLH-----GRVAAAFQVFGNISPFVKFSH 394
DB 873 PDPALRTISE--YARHVMSPGN--RMHPFVPLGAVDPGLGYN--VPALYSD 922
QY 395 FTANQAIQFAFEER-----VHIDLDIMQGLQMFGLFHT-----LASRPGSPRV 441
DB 923 PAARERERARERDLRLKPGFEVKPSLEPLGHVPGGLPPFRHGLALQGPSP-- 979
QY 442 RLTLGLASMEALFATGRKLSDFADTLGLPFECVAEKAAGV-----DPEKLGVTREEA 495
DB 980 -----GLHFPFHPISLG-PLEBERLALMAGPLRDMKSAEFLAERQHA 1023
QY 496 VAVHMLHSLYDVTGSDSNTLWLQRLARVVTYMWEDLSHSGSL-ARFVAIHYSA- 553
DB 1024 ERVAGL-----GNDP-----LARQLMNTVPHHGHSHISHLHLHQODAIHAASAS 1070
QY 554 ---LFPSTL-DASVGEDSP-----ERHYVEQDLISREIRNYA 586
DB 1071 VHPDLIDPLASGSHLRIPYAGTLPNPLLPNPLEHENEVLRLHFAAPYRDLPA 1123

RESULT 10
YK82_MCTTU STANDARD; PRT; 721 AA.
ID YK82_MCTTU
AC Q10690;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 73.6 KDA PROTEIN RV2082.
GN RV2082 OR MTCY49.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborn J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulton J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: SOME, TO M. TUBERCULOSIS RV0029.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: 273966; CAA98194.1; -
 DR Tuberculin; RV2082; -
 KW Hypothetical protein; Complete proteome.
 FT DOMAIN 295 307 ALA-RICH.
 SQ SEQUENCE 721 AA; 73564 MW; 28CAC21029ED0C57 CRC64;

Query Match 5.1%; Score 178; DB 1; Length 721;
 Best Local Similarity 24.7%; Pred. No. 0.0091;
 Matches 96; Conservative 20; Mismatches 125; Indels 148; Gaps 17;

QY 3 PPPPPPLTPYCRRCRPPPHLPP-----PSSPNHFLHYHQDHEAAAAAARRKR 55
 DB 250 PPTAPPTPTSPGTAPMPGPPPGTVSPPLPSAP-----AVGAGPSV 293
 QY 56 PASDMDLPPRRHYTGDLSDVTAAGVY-GSGAPSSASAC--LPALPTQLHQLPPAF 111
 DB 294 PAAGM--PAAAAATAPLSPQSLGQFTTGMTGTAAAGALSGALHAATEPLPP- 350
 QY 112 QHNAPEVDVAHPAHPAQAAGGATATTAWDGIIIRDIIGSSGAAVSTOLLHNRE 171
 DB 351 --APPTPTPTVTPV-----ATATTA--GI----- 372
 QY 172 ITHPCNGLSLLELRSLAADPAPLPPPOQQA-----LHGAPAAAGLTLR 225
 DB 373 ---PHIDSA-----TPSPAPIAPPTIDNASAMPTAPMAVNGPPASAPAPAA 419
 QY 226 PPPPLP---DKRHHHPCCQOOOEHPH-----APQSPH----- 259
 DB 420 PAGPLPAGADLR-----PVTPPPATPTPTGTPTSGAAVTSPSPAAGSLMSPVVKST 474
 QY 260 -----PRAETAAAAAQAQAAAAAKEKEQRKQ-----RDEGHLHLT 301
 DB 475 APATTAQAPSNPTPLSATTAATGTGAAGDTSRAAEQRLRIIDTVAROEGLSW- 532
 QY 302 LLQCAEAVNANDNDHATLLELFAELAT 330
 DB 533 -----AAGLRNGQTTLVTLTDLAS 551

RESULT 11
 ID EXTN_TOBAC STANDARD; PRT; 620 AA.
 AC P13983;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-1992 (Rel. 21, Last annotation update)
 DE EXTENSIN PRECURSOR (CELL WALL HYDROXYPROLINE-RICH GLYCOPROTEIN).
 GN HGPRT3.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. XANTHI; TISSUE=Leaf;
 RX MEDLINE=90128263; PubMed=2612909;
 RA Keller B., Lamb C.J.;
 RT "Specific expression of a novel cell wall hydroxyproline-rich
 RT glycoprotein gene in lateral root initiation.";
 RT Genes Dev. 3:1639-1646(1989).
 RU -1- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN
 CC THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE
 CC MAIN ROOT.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
 CC -1- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN

CC GLYCOSTYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X13885; CAA32090.1; -
 DR PIR: S06733; S06733
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 ?
 FT CHAIN ? 620
 FT REPEAT 70 73 EXTENSIN.
 FT REPEAT 148 151 H-A-P-P.
 FT REPEAT 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.
 FT REPEAT 236 242
 FT REPEAT 205 242
 FT DOMAIN 205 620
 FT DOMAIN 499 600
 FT SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;
 CONTAINS THE SER-PRO(4) REPEATS.
 3 X APPROXIMATE TANDEM REPEATS.

Query Match 4.9%; Score 171; DB 1; Length 620;
 Best Local Similarity 26.3%; Pred. No. 0.018;
 Matches 72; Conservative 10; Mismatches 92; Indels 100; Gaps 12;

QY 2 PPPP---PPPLTPYCRRCRPPPHLPPPPSSPNHFLHYHQDHEAAAAAARRKRAS 58
 DB 348 PPPPPYSPPPP--PSYSPPPPTLPPPPSSP-----PPPSFSPPPPT 389
 QY 59 DMDLPPRRHYTGDLSDVTAAGVSGAPSSAS-----AQLPALPTQLHQLPPAF 112
 DB 390 EOSPPPPPAY-----SPPLAPPTYSPPPTTSPPPPTAQPPLPPPTSPPPAYS 441
 QY 113 HNAPEVDVA---HPAPAQAAGGATATTAWDGIIIRDIIGSSGAAVSTOLLHN 168
 DB 442 PPPPPYSPPPPTYSPPPAYAQ----- 465
 QY 169 VREIITHPCNPGIASLLELRSLAADPAPLPPPOQOHALHGAPAAAGLTLPPPP 228
 DB 466 -----PPPPPTYS-----PPPAYSPPPSP--IYSPPPVOVPLPPTFSPP 506
 QY 229 PLDKRHHHPCCQOOOEHPH--POSPKAPT 261
 DB 507 P---RHILPPPP--HKQPPPTPTTYGDPSPPT 535

RESULT 12
 ID VNDA_PVKA STANDARD; PRT; 1733 AA.
 AC P33483;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE PROBABLE NUCLEAR ANTIGEN.
 OS Pseudorabies virus (strain Kaplan) (PRV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=33703;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021039; PubMed=2171211;
 RA Vieck C., Kozmlik Z., Paces V., Schiim S., Schwyzer M.;
 RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
 RT oriented open reading frame: characterization of their promoter and
 RT enhancer regions.";
 RT Virology 179:365-377(1990).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M34651; AAA47471.1; -.
 DR PIR: B45344; B45344.
 FT DOMAIN 112 117 POLY-THR.
 FT DOMAIN 179 1733 GLY-RICH.
 FT DOMAIN 192 196 POLY-SER.
 FT DOMAIN 271 298 POLY-PRO.
 FT DOMAIN 304 308 POLY-ARG.
 FT DOMAIN 883 889 POLY-GLY.
 FT DOMAIN 1398 1405 POLY-GLY.
 SQ SEQUENCE 1733 AA; 172166 MW; 0C8CDBBE475B5E2 CRC64;

Query Match 4.8%; Score 169.5; DB 1; Length 1733;
 Best Local Similarity 22.0%; Pred. No. 0.07;
 Matches 109; Conservative 27; Mismatches 153; Indels 207; Gaps 20;

QY 2 PPPPPRPL-----TP-TCRCRPPRHLP-----PPSSPNHFLHYLHQLDHOE 45
 DB 70 RPNRRPPQDHHRRTPRARDHNDPRDLPTRTRDHCHRRPPTTTTITKDPHROPDL 129
 QY 46 AAAAMVKKRP---ASDDLPRRRHYTGDSVTAANAAGVSGGAPSSAS----- 94
 DB 130 LPTKLOEDPHLRPTDRPSAKTHNHOD-----PRGGCPRTSSHHNHODPP 178
 QY 95 ----AQLRALP---TQLHQLPRAFHNAPEVDVPAHPA---ANAAGGATASTTA 142
 DB 179 GGGPPSPRRPSTSSSHGSPRTPRRPPRPPRPPRPPRPPRPPRPPRPPRPPRPPRPP 237
 QY 143 WWDGIIRLIIGSSGAASITOLHNVAEII-HCNPGLASLLLELRSLAA-----D 195
 DB 238 -----FSSHENKLFSPHSGEGEG-----DRGTAGGSGGDRDD 270
 QY 196 PARLPPRPOONHALLHAPAAARAGLTLPRLPPRLPDKRHHNPPRCOOQOEHPHPAQ 255
 DB 271 PPPPPRRPPRP-----PLPPPPPP-----PP-----QPPRAGG 301
 QY 256 SP-----KAPTAETAAAAAQAQAAAAAAKERKEORRORDEE- 295
 DB 302 SARRRRRGGCPRRGRRGGKRRRAEGTEAAADAEEEDGDDEDEDRAEGEGRED 361
 QY 296 -----GHLITLTLLOCAEA 309
 DB 362 GGEGRGAGGAGESESESGRAEGAPSAEOQVAGVGLLVVRDGLHLDGPERAAGPA 421
 QY 310 VNADNLDAHQTLLEIAELATPTGTSTQRYAAVFAEAMASRLVSSCLGLVAPLPGSPAA 369
 DB 422 VAAADADDLH-----RVPLVLAGAGRGARPGRPGLHC--APGGAD 461
 QY 370 ARL-----HGRVAA 378
 DB 462 AGLEGGKVEHGRGA 477

RESULT 13
 LA17_YEAST STANDARD; PRT; 633 AA.
 AC Q12446;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROLINE-RICH PROTEIN LASI7.
 GN LASI7 OR YOR181W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Toth-E.A.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RA Hughes B., Polh T.M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO S.POME WSP1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: D78487; BAA11386.1; -.
 DR EMBL: J75089; CAA9390.1; -.
 DR SGD: S0005707; LAS17.
 DR InterPro: IPR000697; RanBPL_WASP.
 DR InterPro: IPR001960; WH1.
 DR InterPro: IPR003124; WH1.
 DR Pfam: PF00568; WH1; 1.
 DR Pfam: PF02205; WH2; 1.
 DR SMART: SM00461; WH1; 1.
 DR SMART: SM00246; WH2; 1.
 FT DOMAIN 185 190 POLY-PRO.
 FT DOMAIN 323 329 POLY-PRO.
 FT DOMAIN 342 348 POLY-PRO.
 FT DOMAIN 352 358 POLY-PRO.
 FT DOMAIN 385 391 POLY-PRO.
 FT DOMAIN 427 431 POLY-PRO.
 FT DOMAIN 470 474 POLY-PRO.
 FT DOMAIN 503 507 POLY-PRO.
 FT DOMAIN 520 526 POLY-PRO.
 SQ SEQUENCE 633 AA; 67571 MW; 448835563AA2645 CRC64;

Query Match 4.8%; Score 168.5; DB 1; Length 633;
 Best Local Similarity 24.5%; Pred. No. 0.026;
 Matches 91; Conservative 36; Mismatches 143; Indels 101; Gaps 18;

QY 3 PPPPPPLTPYCR-----RCPPHLPPPPPSPNHFLHYLHQLDHOEAAAAMVKKRPAS 58
 DB 323 PVPPPPPMTTTEGSGVRLP--APPPP-----RRGAP 354
 QY 59 DMDLPPRRHYTGDSVTAANAAGVSGGAPSSASAOULPALPTOLHQLPRAFOHHPAV 118
 DB 355 P---PPRHHVTSNTLN-----SAG-GNSLLPQATGRGRPAPP-----PPRASPTTPV 400
 QY 119 DVAPHAPAAHA-QAGGEATAP-----TAWDGIIRLIIGSSGAASITOLHNH 169
 DB 401 TMOQNPOQINNSNRPGYOTNSMSSPPPPVTTFTTLPOTATAGORAVLPQ--NT 457
 QY 170 RELIHPNGLASLELRSLAADRAPLPPRPOQHALLHGAAPAAAGLTLPPLPP 229
 DB 458 QA-----PSQATNPV-----APPPPASLGSGQIQSAPSA-----PIPT 494
 QY 230 LPDKRHHNRP---CQOQOEHPHPAPOSKAPTAETAAAAAQAQAAAAAAKERKEE 286
 DB 495 LPSTTSAPRPPPAFLTQPOSGAGAPRPPOMPATSTSGGSPAFETTGDAG---RDA 550
 QY 287 QRKKQKDEGSHLITLTLLOCAEAVNANDNDHQLTLEI-ARLAP-----FGTSTQRYA 340
 DB 551 LASTIGAGGIGAL-----RKVDKSQLDKPSVLLDGRGSSASPPAAGNGTGGGP 603
 QY 341 AYFAEASARL 351
 DB 604 ASLADALNAL 614

```
RESULT 14
DIAP_HUMAN STANDARD; PRT; 1248 AA.
AC 060610; 09UC76;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DIAPHANOUS PROTEIN HOMOLOG 1 (DIAPHANOUS-RELATED FORMIN 1) (DRFL).
GN DIAPH1 OR DIAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98028756; PubMed=9360932;
RA Lynch E.D., Lee M.K., Morrow J.E., Welch P.L., Leon P.E., King M.-C.;
RT "Nonsyndromic deafness DFNA1 associated with mutation of a human
RL homolog of the Drosophila gene diaphanous."
RL Science 278:1315-1318(1997).
RN [2]
RP SEQUENCE OF 218-817 FROM N.A.
RA TISSUE-Ovarian carcinoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Wakamabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NEDD human cDNA sequencing project."
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 727-765 AND 1121-1145.
RC TISSUE=platelet;
RX MEDLINE=9525215; PubMed=7737110;
RA Reinhard M., Gleich K., Abel K., Haffner C., Jarichau T., Hoppe V.,
RA Jockusch B.M., Walter U.;
RT "The proline-rich focal adhesion and microfilament protein VASP is a
RT ligand for profilins."
RL EMBO J. 14:1583-1589(1995).
CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DER PROTEIN COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC SIMILARITY). IN HEARING IT MAY PLAY A ROLE IN THE REGULATION OF
CC ACTIN POLYMERIZATION IN HAIR CELLS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, HEART, PLACENTA, LUNG,
CC KIDNEY, PANCREAS, LIVER, SKELETAL MUSCLE, AND COCHLEA.
CC -1- DOMAIN: DRES ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC RHO-GTP ACTIVATES THE DRES BY DISRUPTING THE GBD-DAD INTERACTION
CC (BY SIMILARITY).
CC -1- DISEASE: DEFECTS IN DIAPH1 ARE A CAUSE OF AUTOSOMAL DOMINANT
CC NONSYNDROMIC SENSORINEURAL DEAFNESS 1 (DFNA1).
CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 3 (FH3) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DRE AUTOREGULATORY DOMAIN (DAD).
CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. DIAPHANOUS
CC SUBFAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN THAT IT SEEMS
CC TO INCLUDE INTRONIC SEQUENCE.
CC -1- DATABASE: NAME=Hereditary hearing loss homepage;
CC NOTE=Gene page;
CC WWW="http://www.uia.ac.be/dnalab/hhh/hhgenes.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF051782; AAC05373.1; -.
DR EMBL: AK023345; BAB14533.1; ALT_SEQ.
DR MIM: 602121; -.
DR MIM: 124900; -.
DR InterPro: IPR003104; FH2.
DR InterPro: IPR002965; P_rich_extensn.
DR Pfam: PF02181; FH2; 1.
DR SMART: SM00498; FH2; 1.
KW Coiled coil; Repeat; Deafness.
FT DOMAIN 63 260 GBD.
FT DOMAIN 157 457 FH3.
FT DOMAIN 460 563 COILED COIL (POTENTIAL).
FT DOMAIN 588 743 FH1 (PRO-RICH).
FT DOMAIN 748 1190 FH2.
FT DOMAIN 1015 1172 COILED COIL (POTENTIAL).
FT DOMAIN 1173 1187 DAD.
FT DOMAIN 1189 1192 ARG/LYS-RICH (BASIC).
FT DOMAIN 804 804 T -> TSKA (IN REF. 2).
FT CONFLICT 1132 1133 RK -> AE (IN REF. 3).
SQ SEQUENCE 1248 AA; 138978 MW; ED1F514CF9A886 CRC64;

Query Match 4.8%; Score 167.5; DB 1; Length 1248;
Best local Similarity 19.6%; Pred. No. 0.062;
Matches 114; Conservative 53; Mismatches 177; Indels 243; Gaps 25;

QY 2 PPPPPPPPLPYCRKCPPLPLP-----PPSSPNHLLHYHQIDQEAAMAAVRRK 55
DB 600 PPPPPPP-----PPPLPGTAISPPPLS----- 625
QY 56 PASDMLPPRRHVTDLSVYTAAMAAAGVGSGAPSSASALP--ALPTQHLQLPFAQ 112
DB 626 --GDATIPPP-----PPLP-----GVGISPS--LPGGTAIP-----PPL 660
QY 113 HNAEVDVPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHPAHP 172
DB 661 PGSAIRIPPPPLP-----GSAG----- 678
QY 173 IHCNGLASLETLRLSLADPAPLPPPOOHALLGAPAAPAGLTPPPPLD 232
DB 679 IPPPPPLPG-----EAGMPPPP-----LPGGGIPPPPPPG 713
QY 233 KRHEHPPCQOQOQEP--HPAPSPKAP-----TAETAA 267
DB 714 GRIPIPPPGMGKPPPPPGFGVPAAPVLPGLTRKLLKPEVQLRRPWSKLVAEELSQ 773
QY 268 -----AAAAAQAAMAAAKERREQRKQRDEGLHLTL 302
DB 774 DCFWTKKEDREFENNELFAKLTLTFSAQTKKDDGGEKKSVQKKVKE----- 824
QY 303 LQCAAAVAADNL-----DDAHTLEIHELATPETSQRYAAATFAEMASRL 351
DB 825 -LKVDSKTAQNLSTFLGFRMPYQEIKNVLEVNE-AVLESMQNLKQMPPEQK 882
QY 352 VSSCLGLVAPLPGPS-----AAARLHGRVAAP-----QVFNGISPFKFSHTAN 398
DB 883 LSLADEYDLDLASEQFGVMQTVRIARRNALILFKLQFSQVFN-----IKPIVSVT 937
QY 399 QAIQEAEREERVHIIDIDIMOGLOWPGLFHTLASRPGPPRVRLTGLGASMEALATGK 458
DB 938 AACCEELRKSEFSNLELTLL-----VGNYNAGSRNAGAFENISFL-CKLRDTKSTDQ 991
QY 459 RL-----SDFADTLGPFEECAVAEKAAGVDEKL 488
DB 992 KMTLLHFLAELCENDYPDVLPDELAVH-EKASRYSAENL 1031

RESULT 15
SEPA_EMENI
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 19:33:15 ; Search time 122.38 Seconds
(without alignments)
798.414 Million cell updates/sec

Title: US-09-265-585C-96

Perfect score: 3503

Sequence: 1 MPPPPPPPLTPYCRRCPP.....KDLCLLASAMPYQVPPCR 668

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3503	100.0	668	10 Q9FUZ7	Q9FUZ7 zea mays (m
2	1865	53.2	653	10 Q9M384	Q9M384 arabidopsis
3	1861	53.1	653	10 Q96304	Q96304 arabidopsis
4	1747	49.9	819	10 Q9AVK4	Q9AVK4 pisum sativ
5	1005.5	28.7	405	10 Q9FH21	Q9FH21 arabidopsis
6	670	19.1	623	10 Q9S759	Q9S759 triticum ae
7	666	19.0	630	10 Q9ST48	Q9ST48 zea mays (m
8	635	18.1	625	10 Q9MB96	Q9MB96 oryza sativ
9	616	17.6	547	10 Q9SRP9	Q9SRP9 arabidopsis
10	602	17.2	493	10 Q9AS97	Q9AS97 oryza sativ
11	601	17.2	511	10 Q9C8Y3	Q9C8Y3 arabidopsis
12	601	17.2	662	10 Q65367	Q65367 arabidopsis
13	594.5	17.0	587	10 Q23642	Q23642 arabidopsis
14	593.5	16.9	587	10 Q23725	Q23725 arabidopsis
15	590.5	16.9	587	10 Q9SLH3	Q9SLH3 arabidopsis
16	590.5	16.9	532	10 Q23724	Q23724 arabidopsis
17	590.5	16.9	532	10 Q9LQ78	Q9LQ78 arabidopsis
18	589.5	16.8	532	10 Q23643	Q23643 arabidopsis
19	574	16.4	617	10 Q9LW09	Q9LW09 oryza sativ

20	572	16.3	523	10 Q9LFE3	Q9LFE3 arabidopsis
21	570.5	16.3	526	10 Q9LPT0	Q9LPT0 arabidopsis
22	562	16.0	490	10 Q9LDL7	Q9LDL7 arabidopsis
23	561	16.0	658	10 Q9CAN3	Q9CAN3 arabidopsis
24	557	15.9	584	10 Q9FL03	Q9FL03 arabidopsis
25	548	15.6	413	10 Q9S7H5	Q9S7H5 arabidopsis
26	544.5	15.5	593	10 Q9SDQ3	Q9SDQ3 arabidopsis
27	538.5	15.4	447	10 Q9ARF6	Q9ARF6 capsella ru
28	536.5	15.3	482	10 Q9LPR8	Q9LPR8 arabidopsis
29	525	15.0	428	10 Q9ZSP2	Q9ZSP2 lycopersico
30	513.5	14.7	445	10 Q9ZWC5	Q9ZWC5 arabidopsis
31	499.5	14.3	306	10 Q9XE53	Q9XE53 arabidopsis
32	498	14.2	352	10 Q9XE51	Q9XE51 arabidopsis
33	470.5	13.4	542	10 Q9SCR0	Q9SCR0 arabidopsis
34	466	13.3	610	10 Q9LRT5	Q9LRT5 arabidopsis
35	458	13.1	1502	10 Q9LNX6	Q9LNX6 arabidopsis
36	457	13.0	483	10 Q9SUF5	Q9SUF5 arabidopsis
37	442	12.6	718	10 Q80933	Q80933 arabidopsis
38	439.5	12.5	583	10 Q9SMB8	Q9SMB8 arabidopsis
39	436	12.4	101	10 Q9AVK3	Q9AVK3 pisum sativ
40	423.5	12.1	375	10 Q23566	Q23566 arabidopsis
41	417.5	11.9	1336	10 Q81074	Q81074 arabidopsis
42	404.5	11.5	287	10 Q9MOM5	Q9MOM5 arabidopsis
43	399.5	11.4	284	10 Q9XE57	Q9XE57 arabidopsis
44	398	11.4	808	10 Q9XE58	Q9XE58 arabidopsis
45	373.5	10.7	582	10 Q9FMN6	Q9FMN6 oryza sativ

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	668 AA.
Q9FUZ7	Q9FUZ7			
AC	Q9FUZ7			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	SCARECROW.			
GN	SCR.			
OS	zea mays (Maize).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;			
OC	Panicoidae; Andropogoneae; Zea.			
OX	NCBI_TaxID=4577;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20407145; PubMed=10948251;			
RA	Lim J., Helariutta Y., Specht C.D., Jung J., Sims L., Bruce W.B.,			
RA	Diem S., Benfey P.N.;			
RT	"Molecular analysis of the SCARECROW gene in maize reveals a common			
RT	basis for radial patterning in diverse meristems."			
RL	Plant Cell 12:1307-1318(2000).			
DR	EMBL: AF263457; AAG13663.1;			
SQ	SEQUENCE 668 AA: 71162 MW; 9287431046B2B621 CRC64;			

Query Match	100.0%	Score 3503;	DB 10;	Length 668;
Best Local Similarity	100.0%;	Pred. No. 2.8e-222;		
Matches 668;	Conservative	0;	Mismatches 0;	Indels 0;
QY	1 MPPPPPPPLTPYCRRCPPPHLP	PPPPSSPNHFLHYLHQDLHQA	AAAAAMRKRPASDM 60	
DB	1 MPPPPPPPLTPYCRRCPPPHLP	PPPPSSPNHFLHYLHQDLHQA	AAAAAMRKRPASDM 60	
QY	61 DLPPRRHVTGDLSDVTA	AAAGVGSGAPSSAQLPALPTQL	QLQLPAPFQHNAPEDV 120	
DB	61 DLPPRRHVTGDLSDVTA	AAAGVGSGAPSSAQLPALPTQL	QLQLPAPFQHNAPEDV 120	
QY	121 PAHPADAAHQAGEATASTA	AVDGIIRDITGSSGGA	AVSITQILHNVELIHCNGL 180	
DB	121 PAHPADAAHQAGEATASTA	AVDGIIRDITGSSGGA	AVSITQILHNVELIHCNGL 180	

```

OY 181 ASLELELRSLLAADPAAPLPPPOPOQHALLHGAPAAAPAGLTLPPPLPDKRHHPP 240
DB 181 ASLELELRSLLAADPAAPLPPPOPOQHALLHGAPAAAPAGLTLPPPLPDKRHHPP 240
OY 241 PQQOQOQEEPRHAPSPAPTAETAAAAAQAQAAAAAERKEEQRKORDEBGLHL 300
DB 241 PQQOQOQEEPRHAPSPAPTAETAAAAAQAQAAAAAERKEEQRKORDEBGLHL 300
OY 301 TLLLOCAEVNNDLDDHOTLLEIAELATPGTSTORVAAVFAEAMARLVSSGLGXA 360
DB 301 TLLLOCAEVNNDLDDHOTLLEIAELATPGTSTORVAAVFAEAMARLVSSGLGXA 360
OY 361 PLPPSSPAARLHGRVAAAFQVFNCSIPVFKSHFTANQAIQEAEREREVHIIDLDIMQ 420
DB 361 PLPPSSPAARLHGRVAAAFQVFNCSIPVFKSHFTANQAIQEAEREREVHIIDLDIMQ 420
OY 421 GLQWGLFHILASRGGPRVRLTGLGSMALERTGKRLSDPADTLGLPEFCVAVEKA 480
DB 421 GLQWGLFHILASRGGPRVRLTGLGSMALERTGKRLSDPADTLGLPEFCVAVEKA 480
OY 481 GNVDEKIGVTRREAVAAVHMLHSLDYVTGSDSNTMLIORLAPKVVYTMVEDLSHSGSF 540
DB 481 GNVDEKIGVTRREAVAAVHMLHSLDYVTGSDSNTMLIORLAPKVVYTMVEDLSHSGSF 540
OY 541 LARFEALHYYSALFDSLDASYGEDSPERHVEEQQLSREIRNVLAAGPARTGVKFS 600
DB 541 LARFEALHYYSALFDSLDASYGEDSPERHVEEQQLSREIRNVLAAGPARTGVKFS 600
OY 601 WREKLAOSGFRAASLAGSAAQAASLLGMFPSSDGYTLVEENGALKLWKDCLLTASAMR 660
DB 601 WREKLAOSGFRAASLAGSAAQAASLLGMFPSSDGYTLVEENGALKLWKDCLLTASAMR 660
OY 661 PLOVPPCR 668
DB 661 PLOVPPCR 668

```

RESULT 2

09M384

PRELIMINARY: PRT: 653 AA.

09M384

09M384

01-OCT-2000 (TREMBlrel. 15, Created)

01-OCT-2000 (TREMBlrel. 15, Last sequence update)

01-JUN-2001 (TREMBlrel. 17, Last annotation update)

SCARECROW1.

F24B22.180.

Arabidopsis thaliana (Mouse-ear cross).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

eurosid II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

SEQUENCE FROM N.A.

Blöcker H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,

Salanoubat N.,

Submitted (NCV-1999) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

EU Arabidopsis sequencing project;

Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

EMBL: AL132957; CAB70996.1;

SEQUENCE 653 AA; 71505 MW; 009A4C48C6DA0616 CRC64;

Query Match 53.2%; Score 1865; DB 10; Length 653;

Best Local Similarity 57.4%; Pred. No. 1.3e-114;

Matches 405; Conservative 66; Mismatches 122; Indels 112; Gaps 18;

OY 4 PPPPPL-----TPYCRCPPLPPPPSSPNHFLHYLQDLHQEAAAAAAYKKRP 56

DB 12 PPHSPPLRTSSSSSSNNNGPP--PPPP-----LVMYRKRL 48

OY 57 ASDM-----DLPPRR--H-VTGDLSDVT-----AAAAAGVGSGAPS---SASAO 96

```

DB 49 ASESSNDVNNSSRRRRSHLSDSNVNTVTPQPSLTAATVSSQNPPLSVCGFSG 108
OY 97 LPALPTQHLQPLPAFOHHADEVDPVPAHPAPAAHQAQAGENTASTAWDGIITDIISSSG 156
DB 109 LPVPPS-----DRGGRNVMSVQPMQ---QSSSSASPTWVDAIIRDLIHS- 154
OY 157 GAAVSITOLHNREIHPNPGSLSLLELRSLLAAPRA-----PLP----- 200
DB 155 -TSVSLPOLQNVNDIIFPCNPMLGALLELRSLMLDLSSSDSPQTFEPPLVQISNN 213
OY 201 -PPPOQOHALHGAAPAAAPAGLTLPPPLPDKRHHPPCCQOQOQEEPRHAPQSPK 258
DB 214 PSPPOQOQOQOQO-----QOQOQHPPPPPIQOQORE-----NSSTD 248
OY 259 APTAETAAAAA-AAQAAAAAERKEEQRKORDEBGLHLTLLLOCAEVNNDLDD 317
DB 249 APQPEVTVATVAVQNTNAELRERKEEIKRQKODEBGLHLTLLLOCAEVNNDLDE 308
OY 318 AHOTLLEIAELATPGTSTORVAAVFAEAMARLVSSGLGXAAPLPPGSPAARLHG-RV 376
DB 309 ANKLELEISLSPYGTSAQRYAAVFAEAMARLVSSGLGXAALP--SRMPQTHSLKM 366
OY 377 AAQOVFNCSIPVFKSHFTANQAIQEAEREREVHIIDLDIMQGLQWGLFHILASRPG 436
DB 367 VSAFQVFNCSIPVFKSHFTANQAIQEAEREDSVHIIDLDIMQGLQWGLFHILASRPG 426
OY 437 GPRVRLTGLGSMALERTGKRLSDPADTLGLPEFCVAVEKAGNVDEKIGVTRREAV 496
DB 427 GPRVRLTGLGSMALERTGKRLSDPADTLGLPEFCVAVEKAGNVDEKIGVTRREAV 486
OY 497 AVHMLHSLDYVTGSDSNTMLIORLAPKVVYTMVEDLSHSGSFARFEALHYYSALFD 556
DB 487 AVHMLHSLDYVTGSDAHTMLIORLAPKVVYTMVEDLSHSGSFARFEALHYYSALFD 546
OY 557 SLDASYGEDSPERHVEEQQLSREIRNVLAAGPARTGVKFSWREKLAOSGFRAASLA 616
DB 547 SLGASGESEERHVEEQQLSREIRNVLAAGPARTGVKFSWREKLAOSGFRAASLA 606
OY 617 GSAAQAASLLGMFPSSDGYTLVEENGALKLWKDCLLTASAMRP 661
DB 607 GNAATQATLLGMFPSSDGYTLVDNGTLKLMKDLTLTASAMRP 651

```

RESULT 3

096304

PRELIMINARY: PRT: 653 AA.

096304

01-FEB-1997 (TREMBlrel. 02, Created)

01-FEB-1997 (TREMBlrel. 02, Last sequence update)

01-JUN-2000 (TREMBlrel. 14, Last annotation update)

SCARECROW1.

Arabidopsis thaliana (Mouse-ear cross).

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

eurosid II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

SEQUENCE FROM N.A.

Di Laurenzio L., Wysocka-Diller J., Malamy J.E., Pysh L.,

Helariutta Y., Freshour G., Hahn M.G., Feldmann K.A., Benfey P.N.,

Cell 0:0-0(0).

EMBL: U62798; AAB06318.1;

SEQUENCE 653 AA; 71535 MW; 502057F2666B161B CRC64;

Query Match 53.1%; Score 1861; DB 10; Length 653;

Best Local Similarity 57.3%; Pred. No. 2.3e-114;

Matches 404; Conservative 66; Mismatches 123; Indels 112; Gaps 18;

OY 4 PPPPPL-----TPYCRCPPLPPPPSSPNHFLHYLQDLHQEAAAAAAYKKRP 56

DB 12 PPHSPPLRTSSSSSSNNNGPP--PPPP-----LVMYRKRL 48

OY 57 ASDM-----DLPPRR--H-VTGDLSDVT-----AAAAAGVGSGAPS---SASAO 96


```
Db 12 PPHSPLRTSSGSSSSNNRGPP--PPPP-----LVMVRKL 48
Qy 57 ASDM-----DLPPRR--H-VTGLSDVT-----AAAAGVSGAPS---SASAQ 96
Db 49 ASESSNDYNNSSRPBRVSHLDSNNVTTPQOPSLTAAATVSSQBPPLSVCGSG 108
Qy 97 LPALPTQHLQRPAPFQHNHAPVDVPAHPARAHAAGGAGTASTTAVNDGIRIITSSG 156
Db 109 LPVPRS-----DRGRNVMSVQPM-----ODSSSSASPTVWDIIRDLHSS- 154
Qy 157 GAAVSTOLHNVEEIIHPCNPGSLASLELRSLADPA---PLP----- 200
Db 155 -TSVSTPQLIGNVDIIFPCNPNGALLFVRLSLMLDSSSDPSPQFPEPLYSNN 213
Qy 201 --PPPOQAHLHGAPAAAGLTLPPRPLPDKRRHHPPCQOQOQEEHPAPQSPK 258
Db 214 PSPPOOQOHH-----OQOQOHHKPPPPYIQOQERE---NSTD 248
Qy 259 APTAEETAAAA--AAOAAAAAAKERKEQRRKQDEEGHLLTLTQCAEVNADLDD 317
Db 249 APPOPEVTATVPVAVQNTAEALRERKEETKROKODEGHLTLTQCAEASADNLEE 308
Qy 318 AHQTLLEIAELATPEGTSTORVAAVFAEASARLVSSCLGLYAPLPGPSAARLHG-RV 376
Db 309 ANKLLLEISQSTPYGTSAQVAAVFEASARSRLNSCLGIYALP--SRMMQOTHSKLM 366
Qy 377 AAAQVQNGISPEYKESHFTANQAIQEAEREERVHIIDIMOGLOWPGLFHTLASRPG 436
Db 367 VSAQVQNGISPLVKFHSFTANQAIQEAEREKDSVHIIDIMOGLOWPGLFHTLASRPG 426
Qy 437 GPRVRLTGLGASMEALATGKRLSDPADTLGLPEFCAVAEKGANVDPKLGVTREAV 496
Db 427 GPRVRLTGLGTSMEALQATGKRLSDPTDKLGPEFCAVAEKGANVDPKLGVTREAV 486
Qy 497 AVHMLHSLYDVTGSDSNTLMLIQRAPKYVTVEODLSHSGFLARFEVAIHYSALFD 556
Db 487 AVHMLHSLYDVTGSDAHTLMLIQRAPKYVTVEODLSHSGFLARFEVAIHYSALFD 546
Qy 557 SLDSYGEDSEERHVEQQLSREIRNVAVGAPRTGVDYKFGSWREKLAQSGFRASLA 616
Db 547 SLGASVEESEERHVEQQLSREIRNVAVGAPRTGVDYKFGSWREKLAQSGFRASLA 606
Qy 617 GSAAQAQSLILGMFPDGYTLVEENGALTKGMKDLCLLTASAMP 661
Db 607 GNAATQATLLIGMFPDGYTLVDNGTLTKGMKDLCLLTASAMP 651

RESULT 4
Q9AVK4 PRELIMINARY: PRT: 819 AA.
AC Q9AVK4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE SCARECROW GENE REGULATOR-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=CV. ALASKA;
RA Sassa N., Matsushita Y., Nakamura T., Nyunoya H.;
RT "Molecular characterization and in situ expression pattern of pea
RT SCARECROW gene.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB048713; BAB39155.1;
SEQUENCE 819 AA; 90372 MW; 41867BD6DC72ADFA CRC64;
```

Query Match 49.9%; Score 1747; DB 10; Length 819;

```
Best Local Similarity 62.9%; Pred. No. 9,5e-107;
Matches 361; Conservative 64; Mismatches 91; Indels 58; Gaps 8;
Qy 135 EATASTTAWVDGIIRDIILSSGGAASITQLIHNREIHPNPGSLASLELRSLTAA 194
Db 245 EETSATTNMDIGLKDILHTSN--SVSIPQLNNVREIIYPCNPALVLEHRLR--LIT 300
Qy 195 DPAPLPPPPQ---PQOHALHGAPAA-----PAGTLPPPPPLPD 232
Db 301 EPNTQPEKRRNSTEQQGVNNGVNLAAASNVSVKLMNRVDVYPLTSHFSDSTLLN 360
Qy 233 KRHEH-----PPPCQOQOQEEHPAPQSPK--APTAEETAA 267
Db 361 QONQNMFPNMGATQINNPNPVSIVTLPLSOPLTQDOQOHLQOHPEDLAP-----AT 415
Qy 268 AAAAAAATAAAKERKEQRRKQDEEGHLLTLTQCAEVNADNDDAHQTLLEIAE 327
Db 416 TTTTTSABELARKKKEELKEOKKDEBHLTLTQCAEASVENLEQAKMLLEISQ 475
Qy 328 LATPEGTSTORVAAVFAEASARLVSSCLGLYAPLPGPSAARLHGAVAAQVQNGIS 387
Db 476 LTPPEGTSAQVAAVFEASARSRLNSCLGIYATLTPVSSHTPH--NQKVASAQVQNGIS 533
Qy 388 PFYKESHFTANQAIQEAEREERVHIIDIMOGLOWPGLFHTLASRPGPYVRLTGLG 447
Db 534 PFYKESHFTANQAIQEAEREERVHIIDIMOGLOWPGLFHTLASRPGPYVRLTGLG 593
Qy 448 ASMEALATGKRLSDPADTLGLPEFCAVAEKGANVDPKLGVTREAVAMVHLHSLYD 507
Db 594 TSMETLEATGKRLSDPANKLGLPEFFPVAEKVGNIDVEKLVNSSEAVAMVHLHSLYD 653
Qy 508 VTGSDSNTLMLIQRAPKYVTVEODLSHSGFLARFEVAIHYSALFDSYGEDSP 567
Db 654 VTGSDSNTLMLIQRAPKYVTVEODLSHSGFLARFEVAIHYSALFDSYGESE 713
Qy 568 ERHVVEQQLSREIRNVAVGAPRTGVDYKFGSWREKLAQSGFRASLAQAAASILL 627
Db 714 ERHVVEQQLSREIRNVAVGAPRTGVDYKFGSWREKLAQSGFRASLAQAAASILL 773
Qy 628 GMFPDGYTLVEENGALTKGMKDLCLLTASAMP 661
Db 774 GMFPDGYTLVDNGTLTKGMKDLCLLTASAMP 807

RESULT 5
Q9FH21 PRELIMINARY: PRT: 405 AA.
AC Q9FH21;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE SCARECROW GENE REGULATOR-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RE SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA MEDLINE-99397451; PubMed-10470850;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.
RT Sequence features of the regions of 1,011,550 bp covered by seventeen
RT pl and TAC clones.";
RL DNA Res. 6:183-195(1999).
DR EMBL: AB017067; BAB08425.1;
SEQUENCE 405 AA; 44893 MW; 6809D9EEBEB0A8F6 CRC64;
```

Query Match 28.7%; Score 1005.5; DB 10; Length 405;
Best Local Similarity 53.2%; Pred. No. 2.1e-58;

Matches 207: Conservative 65; Mismatches 106; Indels 11; Gaps 6;

OY 278 AAAREKE--EQRKKQDEBGLHLLTLLOCAEAVNADNDDAHQTLLLELAETATPGTGS 335
 Db 17 SSARRLEEFPEETLENDGAAAIKLLSLLOCAEYVAADHLREASTLLSEISEICSPGSS 76
 OY 336 TORAAAYEAEMASRLVSSCL-GYAPLPFGSPAAARLHGAVAAAFQVFNGISPFVFSH 394
 Db 77 PERVAATYAOALQTRVIVSSYLSGACSPLE-KPLTVVOSQKIPISALQTVSVSLIKFSH 135
 OY 395 FTANAOIAEAFEREERVIHIDLMIGLQWPGLEFHILASRPGPPRYRLTGLGASMALE 454
 Db 136 FTANAOIAFOALDGEDSVHIIIDLVMOGLQWPFALFHILASPRKLRISIRITGFGSSDILA 195
 OY 455 ATGRKLSFADTGLPPEFCAVAKAGN-VDEPKLGYTREAVAVHMLHSIVDYGTSGDS 513
 Db 196 STGRRLAFASLLNPEFPIEGITIGNLIDPSQATRGGAAYVHHMQRKLDVYGNL 255
 OY 514 NTLMLIORLAKPVYTWVEODLSH--SGSFARFEVAIHYYSALEPDSIADSYGEDSPERHV 571
 Db 256 ETELELRKLNLTIVVQELSTYDGGSFGRFVEALHYYSALFDALDGDGEGESGERT 315
 OY 572 VEQQLSREIRNVLAAGVPARTGVKFGSWREKLAQSGFPRAASLAGSAAQAQSLLGMP 631
 Db 316 VEOYLGTGEIRNIYAHGSGRR---KRMKMEELSRVGFVSLSRGNPATQACILGLMP 371
 OY 632 SDGYTLVEENGALKLGWKDLCCLLTASAMR 660
 Db 372 WNGTYLVEENGTLRGWKDSLTLTASAMK 400

RESULT 6
 O9ST59 PRELIMINARY: PRT: 623 AA.

AC O9ST59: 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE GIBBERELIN RESPONSE MODULATOR.
 GN RHT-DIA.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OC NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9347734; PubMed=10421366;
 RA Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,
 RA Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,
 RA Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;
 RT "Green revolution" genes encode mutant gibberellin response
 RT modulators.";
 RL Nature 400:256-261(1999).
 DR EMBL: AJ242531; CAB5155.1; -.
 SQ SEQUENCE 623 AA; 65337 MW; F2BAC34996D0A84F CRC64;

Query Match 19.1%; Score 670; DB 10; Length 623;
 Best Local Similarity 30.9%; Pred. No. 4.2e-36;
 Matches 202; Conservative 77; Mismatches 262; Indels 112; Gaps 19;

OY 79 AAAAGVGSGAPSSASAOPLPQLHQLPFAFHNAPEVNVPAHAPAAHQAQSGATA 138
 Db 8 AGSGGGGGGKSSSEDKMVS-----AAAGGCEFDV-----ELLALGYVRA 50
 OY 139 STTAVWDIIRDI--IGSSG--GAAYSTIQLIHNVRETIHPCNPGIASILLRLSL 192
 Db 51 SDMAVAVAKLBDLEAMMGVGAAPDDSFATHLATDVHYNPTDLSVSEML-SEL 109
 OY 193 AADPAPLPPO-----PQHALHNGAPAAAPAGLTLP---PP 227
 Db 110 NAPPPLPAPOLNASTSVTGGGVFDLPSPVSDSSSIYALRPISPGATAPADLSA 169

OY 228 PRLPDKRRHEHPPPCQOQOQEEPPHAPQSPKAPTAETAAAAAQAAMAAKERRKEQ 287
 Db 170 DSVYRDKRRKRTGSSSTSSSSSSSSISGGAKSSVVEAPRVAAMAAATPALPVVV----- 224
 OY 288 RKRQDEBGLHLLTLLOCAEAVNADNDDAHQTLLLELAETATPGTSTORVAAVFAEM 347
 Db 225 --VDTQAGIRLVHALLCAEAVQOENLSAALVKQIFLLAASOGAMRKVAAVGEAL 282
 OY 348 SARLVSSCGIAPLPFGS-----PAAARLHGAVAAAFQVFNGISPFVFSHTANAOIAE 403
 Db 283 ARVFR-----FRPDSSLLDAARADLLH-----FYESCYLFAHETAQALIE 330
 OY 404 AFEREERVIHIDLMIGLQWPGLEFHILASRPGPPRYRLTGLGASMALE 459
 Db 331 AFAGCNRYVVDGKQGNQWALLQALALRRGSGSPFLITGVGPPQDEIDALQOVWK 390
 OY 460 LSDPADTGLPPEFCVA-----EKAGVDEPKLGYTREAVAVHMLH--HSL 505
 Db 391 LAQFAHTIRVDYRGLVATLADLEPFMLQEGEDPNE---BPEYLAIVSVEMHRL 446
 OY 506 YDVTGSDSNTMLIORLAKPVYTWVEODLSH--SGSFARFEVAIHYYSALEPDSIADSYGE 564
 Db 447 LAQPGALEKVLGTVAVRPRIVTVEQENHNSGTFLORETSLHYSTMPDLEGGSSG 506
 OY 565 DSPER-----HVEOOLLSREIRNVLAAGVPART-GDYKFGSWMREKLAQ 608
 Db 507 GGPSEVSSGAAAPAAAGTDQVMSFVYLGROICNVYACGARTERHETLGWMRRRLGNA 566
 OY 609 GFRASLAGSAAQAQSLLGMP-SDGYTLVEENGALKLGWKDLCCLLTASAMR 660
 Db 567 GFETVHLSGNAYKQASTLLALFAGDGVYKEKEGCTLGWHTRLIATSAWR 619

RESULT 7
 O9ST48 PRELIMINARY: PRT: 630 AA.

AC O9ST48: 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE GIBBERELIN RESPONSE MODULATOR (FRAGMENT).
 GN Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 OC NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9347734; PubMed=10421366;
 RA Peng J., Richards D.E., Hartley N.M., Murphy G.P., Devos K.M.,
 RA Flintham J.E., Beales J., Fish L.J., Worland A.J., Pelica F.,
 RA Sudhakar D., Christou P., Snape J.W., Gale M.D., Harberd N.P.;
 RT "Green revolution" genes encode mutant gibberellin response
 RT modulators.";
 RL Nature 400:256-261(1999).
 DR EMBL: AJ242530; CAB51557.1; -.
 FT NON_TER 630
 SQ SEQUENCE 630 AA; 66028 MW; 3D56851726C51042 CRC64;

Query Match 19.0%; Score 666; DB 10; Length 630;
 Best Local Similarity 31.9%; Pred. No. 7.7e-36;
 Matches 209; Conservative 80; Mismatches 253; Indels 114; Gaps 21;

OY 71 GDISDVTAAAGVGSGAPSSASAOPLPQLHQLPFAFHNAPEVNVPAHAPAAH 130
 Db 16 GSKDKMMAAAGAG-----EQEEDVD-----ELLA 42
 OY 131 QAGGEATASTTAVWDIIRDI-----IGSSGAAYSTI--LIHNVRETIHPCNPGIASL 183
 Db 43 ALGYKRVSSDMAVAVAKLBDLEAMMGVGAAGATADDGFSHLATDVHYNPTDLSV 102

QY	184	LEHLRLSLIADPAPLRP--PPQPGQHALLNGAPAAAGLTLPPRPPLPYKRRHEH----	238
Db	103	VESML-SELNPPRLPAPLPAPPLASTSTSTVTSGAAAAGYFDDLPPAVDSSSTYALKP	161
QY	239	-PPCCQOQOQOEPPAPQSPKAPPAEETA-----	272
Db	162	IPSPYAAPASAD---PSYDSNAEPKRMKTTGGGSTSSSSSSSSSDGCGTSSVEAPAPAT	218
QY	273	QAAAAAAKKERKEBQRKKORDEGLHLTLTLLOCAEAVNADNLDDAHQTLLETAELATPP	332
Db	219	QASAAANPAPPVVVVDLTQ--EAGIRLVHALLCAEAVQOENFSAEALVKQIPMLASSQ	276
QY	333	GTSQGRVAAVYAAEAMASARLVSSCGIATLAPLPGSPPAARLHGVAAAFQY-FNGISPEFK	391
Db	277	GGAMKVAAYGEGELARR-----VYRFRP--PDSSLDAAPADLLAHAFYESCYYLK	327
QY	392	FSHFPTANOAIQEAEREERERHVIDLIDMOGLQWPGLEPHILASRPGCPYPRYLTGLCA--	448
Db	328	FAHFPTANOAILLEAFAGCGRVHVDFGIKQGMQWPAALLQALALRGCGPPSRLLGVGPPQ	387
QY	449	-SMELLETGKRISDFADITGLPPEFPCA-VAEKAGNDPEKLV-----TREANAVAH	500
Db	388	DETALQOVGKRLQAFHTIRVDEQYKGLVAATLADLEPMLQPEGDDTDEDEVIANVS	447
QY	501	LH--HSLVDYVDSQSNFTLMLIORLAPKVTVMVEODLSH-SGSFLARFVEAITHYSALFDS	557
Db	448	VPELHRLLAQGALEKVLGYTRAVRPRIYVDEQANHNSTFLDRTESLHYSTMFDS	507
QY	558	L-----DASYGDSPERHVVEEQLLSREIRNVLAVGGPART-GDYKFGSWREKL	605
Db	508	LEGAGAGSGQSTDSPPAAGGTDQDMSEVYLGRIQCVVACEGEAFETREHETLQGMWSRL	567
QY	606	AQSGFRASLAGSMAAQAQSLILGMFP-SDGYTYEENGAKLIGKDKCILTLTASAMR	660
Db	568	GGSGAPVHLDSNAYKQASTLLALFAGDGYRVEEKDCGLTLGWHTRPLTATSAMR	623
RESULT	8		
Q9MB96			
ID	Q9MB96	PRELIMINARY;	PRT: 625 AA.
AC	Q9MB96:		
DT	01-OCT-2000	(TREMBLrel. 15, Created)	
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)	
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)	
DE	OSGAI.		
GN	OSGAI.		
OS	Oryza sativa (Rice).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;		
OC	Ehharctioideae; Oryzoae; Oryza.		
OX	NCBI_TaxID=4530;		
OX	11		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. NIPPONBARE;		
RX	MEDLINE=20179680; PubMed=10713441;		
RA	Ogawa M., Kusano T., Katsumi M., Sano H.;		
RT	"Rice gibberellin-insensitive gene homolog, OSGAI, encodes a nuclear-		
RT	localized protein capable of gene activation at transcriptional		
RL	Gene 245:21-29(2000).		
SR	EMBL, AB030956; BAA90749.1; -		
SO	SEQUENCE 625 AA; 65406 MW; 034FF02719D42E97 CRC64;		

	Query Match	Similarity	18.1%	Score 635	DB 10	Length 625
Best Local	Similarity	32.6%	Pred. No. 8	4e-34		
Matches 186	Conservative	68	Mismatches 233	Indels 84	Gaps 16	
153	GGGGGAAGVSTQLHNVRREIHPGNGGLASLLELRISLLAADPAVLPDPPOQGHALLH	212				
db	72	GVSAGCAADDDGVSHLADTVHVNPSDDLSWVESMSLEINAPLP-PIPPAPPAARAS	130			

QY 213 GA -----PAAAPAG-----LILP-----PPPLPDKRRHNEHPPCOQOOOE 248

Db 131 STVTGGGGSGFFELPAAADSSSTVALRPISLPVATADPSAADSPARTKMRGTGGGSTS 190

QY 249 EPHPAQSPKAPTAEATAAAAAAQAQAAAAAAKEREKEDQRKQRDEGLHLTLTLQCAE 308

Db 191 SSSSSSSSLGGGASGSGVVEAAPPTQGAANAAPVAVVVYVDIOEAGIRLVHALLACAE 250

QY 309 AVNAQNLDDAHQTLLEIAELATPEGTSTQRYAAVPAEAMSAKLVSSCLGLYAPLPQSPGA 368

Db 251 AVQGENFPAALAEALVYQIDPTLAASGGGAMRKVAAYFGELARR-----YVRFPADSTL 303

QY 369 -----AARLHGAVAAAFQVENGISPFYKFSHETANQAIQAEFEERERHVTIDLDIMGLQ 423

Db 304 LDAAPFADLLH-----FYESCYYLTFAPHTANQALILEAFAGCHRVAVVDFGIQCMQ 356

QY 424 WPGFLHILASRPGSPRRVRLTGLA-----SMALFATGKRLSDPFDLTGLRPEPCAV--- 476

Db 357 WFALLQALALRPGGSPRRLTGVGPPQDEFDALQGVCKLAQFAHTTRVDFQYRGVAA 416

QY 477 -----AAKAGNVDPPEKIGVTRREAVAHVHMLHSLLDYTGSDNSTLMLQRLA 523

Db 417 TLADLEPFLPQREGADANDEBEVIAVN-----SVFELHRLLAQ-PGLAEKLVGVHAYR 470

QY 524 PKVVTWVEDDUSH-SGSFLARFEVAIHYSALPDSLD--ASYGDSPER-----HV 571

Db 471 PRIVTVEEAEAHNNGSFLDRFTESLIHYSTWFDSLEGSSGQALSPAPAGGGGGTDOV 530

QY 572 VQOQLSREIRVULVAVGGART-GDVKFGSWMEKLAQSGFRAASLAGSAAQAQSLILGMF 630

Db 531 MSEVYLGRICVAVCEGERTERHETLIGQWNRNIGRAGFEVHIGSNAYKQASTLLALF 590

QY 631 P-SDGYTVEENGALKLGMKDCILTLTASAR 660

Db 591 AGGDGRYVEKEGCLTLGWHTRPLLATSAMR 621

RESULT 9

Q9SRP9

AC Q9SRP9 PRELIMINARY; PRT; 547 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE RGAL-LIKE PROTEIN.

GN T21P5.13.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RP 11

RA SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA;

RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,

RA Romling C.M., Koo H., Fujii C.Y., Utechtack T.R., Barnstead M.E.,

RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;

RT "Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence.";

RL Submitted (0CCT-1999) to the EMBL/Genbank/DBJ databases.

EMBL; AC009895; AF01590.1; --

SQ SEQUENCE 547 AA; 60493 MW; C4D1BD5951D95634 CRC64;

```

Query Match      17.6%  Score 616;  DB 10;  Length 547;
Best Local Similarity 34.8%  Pred. No. 1.3e-32;
Matches 152;  Conservative 82;  Mismatches 175;  Indels 28;  Gaps 11;

QY  242 CQOQOQEEHPQSPKPTAETETAAAAAQAQAAAAAKERKEQRRK---QRDEGL 297
      | : : : | | | : : : | : | : | : | : | : | : | : | : | : |
DB  119 CYDBREYDRAIRPGLSAPFKKEEYVDEEASSKRIILGSMCESDSTSYVLVDQELGV 178
      | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY  238 HLLTLLLOCAEAVNADNDDAHQTLLLETAELATPPTGSTQRYAAFAEAMSARL--VSS 354
      | : | : : : | : : | : : | : : | : : | : : | : : | : : | : : |

```

```

Db 179 RLVAHALVACAEAIHOENINLADALVKRVGTLGASGAGMKVATYFAQALARRIYRDYTA 238
OY 355 CUGIYAPLPSPGAARLHGRVAAAFQVFNIGISPVKFSHFTANOAIQAEFEREREVHII 414
Db 239 EFDVCAAVNPSFEVLEHNI-----FYESCPIYKRAHHTANQALIEATYARRHAVI 289
OY 415 DIDIMQGLQWPGLEPHILASRPGGPPRVRLTIGG---ASMALERTGKRLSDPADITLGLP 470
Db 290 DGLNGMQMWPALMQALALRPGGPPSFRITGIGPQTEINSQSLQGLMKLAQFQNMNGVE 349
OY 471 EEECAV-AEKAGNVDPKLVG-TREAAVAVHHL--HSLYDYGSDSTMLLIQRLAKV 526
Db 350 FEEKGLAEISLSDLEPEMEFTREPESETLVVNSVFELHRLIARSGSIEKILNTVAIKPSI 409
OY 527 VMVEBODLSHSG-SFLARFEVAIHYSALFSDLSAYGEDSPERHVVEQQLSREIRVYL 585
Db 410 VAVVEQGANHNIGIVFLDFRNEALHYSSLFDSLEDSTYSLPQDR-VMSEVYLQGLQILNV 468
OY 586 AVGPARPGDVKFGS-WREKLAQSGFRAASLAGSAAQASILLGNFSPS-DGYTLVEENGA 643
Db 469 AAEQSDRYVERHETAAQMRIRKMSAGFDPIHLGSSAFKQASMLSLYATGDGYRVEENDGC 528
OY 644 IKGKMDCLITLTSAMR 660
Db 529 LMTGMOTRPLTTTSAMK 545

```

RESULT 10

```

OY 09AS97 PRELIMINARY: PRT: 493 AA.
AC 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE PUTATIVE OSGAL.
GN P0707D10.30.
OS Oryza sativa (Rice)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPOBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RA "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:p0707D10."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002910; BAB40172.1; to the EMBL/GenBank/DBJ databases.
SQ SEQUENCE 493 AA; 51864 MW; 5FD8670EB8899492 CRC64;

```

Query Match 17.2%; Score 602; DB 10; Length 493;
 Best Local Similarity 35.2%; Pred. No. 9, 1e-37;
 Matches 171; Conservative 61; Mismatches 192; Indels 62; Gaps 16;

```

OY 195 DPAPLPPOQOQHALLHGAPAAVAGTLRPPRLDPKRHNHPPPCQOQOQDEPPHAP 254
Db 2 DTFPPQWMDP-----AASSGLDAGFLPPAAVAPDDGVGYDP-----PAG 43
OY 255 OSPKAPTEETRAA-AAAAAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 312
Db 44 ADVDAALPPEFAAAPPAPCAPOAAVAVLMMREER-----EVAQIRLVHLLMSCAGATGA 97
OY 313 DN-----LDDAHOITLEIATLPTGSTQORVAAVFAEAMRSARLVSSCLGAPLP 365
Db 98 GDHALASQLADSHALAAVS-AASGIG---RVAHHTTALSRLL-----PSPVA 144
OY 366 SPAARLRGVAAAFQVFNIGISPVKFSHFTANOAIQAEFEREREVHII DIDIMQGLQW 425
Db 145 PPTTDAEH---AFLYHNHYEACPYLKFAHFTANQALIEAFHGCCHVHYVDESLMGOGLMP 201
OY 426 GLFHILASRPGGPPRVRLTIGG---ASMALERTGKRLSDPADITLGLPPEFCVAAEKA- 480

```

```

Db 202 ALIQALALRPGGPPFLRTITGIPSPPTGRDELADVGRLADLARSVVFESFGVAANSL 261
OY 481 GNVDPKLVGTREAVNA---VHMLHSLYDVTSSD--SNLMLLIQRLAKVYIMVQDLS 535
Db 262 DEVRPWWLQIAPGEAAVNSVLIQLHRLGDPADQAPIDAVLDCVAVSRKIFVIEIQEAD 321
OY 536 HSGS-FLARFEVAIHYSALFSDLSAYGEDSPERHVVEQQLSREIRVLAAGPA-RT 593
Db 322 HNTGFLDRTEALFYTSANFSDLSAASGAGNMAE-ATLQREICDVCCEGAAR 380
OY 594 GDVKGFWREKLAQSGFRAASLAGSAAQASILLGNFSPDGYTLVEENGALKMGKDLCL 653
Db 381 RHEPLSRWRDLRTAGISAVPLGSMALRQARMLVGLFSGEGHVEBADCLITLGMHGRPL 440
OY 654 LTASAW 659
Db 441 FSASAW 446

```

RESULT 11

```

OY 09C8Y3 PRELIMINARY: PRT: 511 AA.
AC 09C8Y3;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE GIBBERELLIN REGULATORY PROTEIN, PUTATIVE.
GN T27F4.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri J., Boman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetska I., Kutz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Martelli A.,
RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uiterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408:816-820(2000).
DR EMBL: AC020665; AAG52171.1;
SQ SEQUENCE 511 AA; 56754 MW; 1E60071697C92A9F CRC64;

```

Query Match 17.2%; Score 601; DB 10; Length 511;
 Best Local Similarity 34.6%; Pred. No. 1, 1e-31;
 Matches 146; Conservative 70; Mismatches 150; Indels 56; Gaps 10;

```

OY 273 QAAAAAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 313
Db 107 RAIPGSAYVPRDEHVTIRSKRTIESELSSTRSVVVLDSEQETGVRLVHALLACAAVQON 166
OY 314 NDDAHOITLEIATLPTGSTQORVAAVFAEAMRSARLVSSCLGAPLPSPGAARLH 373
Db 167 NLKLADALVYKHVGLLASGAGARKKAVYTFAGLARIT-----RIY 208

```



```

Db      410 SPTFAVAVNSVFEHLKLRGPGIEKVLGVYKQIKPVLFYVEQESNHNCPVFLDREFTES 469
      : |||||: : | | | | : : : : : | | | | : | | | | :
Qy      548 IHYSALFDSLDASYGEDSPERHVVQOLLSEIRNVLAAGSPART-GDVYFGSGWREKLA 606
      : |||||: : | | | | : : : : : | | | | : | | | | :
Db      470 LHYSTLFDSLE---GVPSQDKVMSEVYLGKQICNLVACGPGPRVERHETLTSQWGNRFG 526
      : |||||: : | | | | : : : : : | | | | : | | | | :
Qy      607 OSGFRAASLAGSAAQAASLLGMFPPS-DGYTLVEENGALKGMKDLCLLTASAMR 660
      : |||||: : | | | | : : : : : | | | | : | | | | :
Db      527 SSGLAFAHLGSAVAFKQASMLLSVFNSSGQYRVESNCGMLGMHTRPLTTTSAMK 581
      : |||||: : | | | | : : : : : | | | | : | | | | :

RESULT 14
ID      023725      PRELIMINARY;      PRT;      587 AA.
AC      023725:
DT      01-JAN-1998 (Tremblrel. 05, Created)
DT      01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT      01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE      GRS PROTEIN.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=COLUMBIA;
RA      Peng J., Carol P., Richards D.E., King K.E., Cowling R.J.,
RA      Murphy G.P., Harberd N.P.;
RL      EMBL; Y15194; CAA75493.1; -
DR      Mendeil; 24071; Arath; 3051.24071.
SQ      SEQUENCE 587 AA; 64006 MW; F6FGC738EE7DCA9 CRC64;

Query Match      16.9%; Score 593.5; DB 10; Length 587;
Best Local Similarity 31.8%; Pred. No. 4.1e-31;
Matches 170; Conservative 83; Mismatches 209; Indels 73; Gaps 18;

Qy      165 LIHNVREIHPNCPG-LASLELRRLSLAADPAPLPPOQOQHALLHGAPAAAPAGL- 222
      : | | | | : : | | | | : | | | | : | | | | :
Db      81 LSHLATDVH-YNPSELYSWLDNML-----SELNPPPLP-----ASSNGLD 120
      : | | | | : : | | | | : | | | | : | | | | :
Qy      223 -TLPP-----PLPDKRREHPPPCQOQOQEEHPAPQSPKAPTAETAA- 268
      : | | | | : : | | | | : | | | | : | | | | :
Db      121 PVLSPETICGPASDYDLEVPGNALYQFPALDISSSSNNQKRLKSCSPDSMTSTST 180
      : | | | | : : | | | | : | | | | : | | | | :
Qy      269 -----AAAAQAAAAAAKKEKEQKQKQDEEGHLITLLLOCAEAVNADNLDA 318
      : | | | | : : | | | | : | | | | : | | | | :
Db      181 GTQIGVIGTIVTTTTTTTAAESTRSVILLVDSQENGVRVLAHLMACAIAIOQNNLTJA 240
      : | | | | : : | | | | : | | | | : | | | | :
Qy      319 HQTLEIAELATPFGSTORVAAYFAEAMSAKLVSCLGLVAPLPQSPAAARLHGVA 378
      : | | | | : : | | | | : | | | | : | | | | :
Db      241 EALVQKQIGCLAVSQAGAMRKVATVYFAELAR-----YRLSPQN-----QIDHCLSD 289
      : | | | | : : | | | | : | | | | : | | | | :
Qy      379 AFQV-FNGISPFKFSHTANAOIAEAFEREERVHIIDIMOGLOWPGLPHILLASRPG 437
      : | | | | : : | | | | : | | | | : | | | | :
Db      290 TLOMHFEYTCRYLKFAPHTANQALILEAFEGKKRVHVIDFSMNOGLQWALMALALREG 349
      : | | | | : : | | | | : | | | | : | | | | :
Qy      438 PPRVRLTGLGA---SMEALATGKRLSDPADTLGLPFEFCA-VAEKAGVNDPEKLG- 490
      : | | | | : : | | | | : | | | | : | | | | :
Db      350 PPTFRITGIGPPAPDNDSHLHEVGCKLAQLAELAHVEEYRGFAVNSLADIASMLRLP 409
      : | | | | : : | | | | : | | | | : | | | | :
Qy      491 TRREAVAVHMLH--HSLVDYTGSDSNTLMLQRLAPKVVMVWEODLSHSGS-FLARFEVA 547
      : | | | | : : | | | | : | | | | : | | | | :
Db      410 SPTFAVAVNSVFEHLKLRGPGIEKVLGVYKQIKPVLFYVEQESNHNCPVFLDREFTES 469
      : | | | | : : | | | | : | | | | : | | | | :
Qy      548 IHYSALFDSLDASYGEDSPERHVVQOLLSEIRNVLAAGSPART-GDVYFGSGWREKLA 606
      : | | | | : : | | | | : | | | | : | | | | :
Db      470 LHYSTLFDSLE---GVPSQDKVMSEVYLGKQICNLVACGPGPRVERHETLTSQWGNRFG 526
      : | | | | : : | | | | : | | | | : | | | | :
Qy      607 OSGFRAASLAGSAAQAASLLGMFPPS-DGYTLVEENGALKGMKDLCLLTASAMR 660
      : | | | | : : | | | | : | | | | : | | | | :

```

```

Db      527 SSGLAFAHLGSAVAFKQASMLLSVFNSSGQYRVESNCGMLGMHTRPLTTTSAMK 581
      : |||||: : | | | | : : : : : | | | | : | | | | :

RESULT 15
ID      09SLH3      PRELIMINARY;      PRT;      587 AA.
AC      09SLH3:
DT      01-MAY-2000 (Tremblrel. 13, Created)
DT      01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT      01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE      PUTATIVE RGAL1, GIBBERELLIN RESPONSE MODULATION PROTEIN.
GN      ATG01570.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CV. COLUMBIA;
RA      MEDLINE=20083487; PubMed=10617197;
RA      Lin X., Kaul S., Rounsley S.D., Shea T.P., Beilto M.-I., Town C.D.,
RA      Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA      Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA      Cronin L.A., Shen M., VanAken S.F., Umayam L., Tallon L.J., Gill J.E.,
RA      Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA      Coppenhaver G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA      Salzberg S.L., Fraser C.M., Venter J.C.;
RT      "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL      Nature 402:761-768(1999).
DR      EMBL; AC005560; AAC67333.1; -
SQ      SEQUENCE 587 AA; 64035 MW; FC92E7F9408072AA CRC64;

Query Match      16.9%; Score 593.5; DB 10; Length 587;
Best Local Similarity 31.8%; Pred. No. 4.1e-31;
Matches 170; Conservative 83; Mismatches 209; Indels 73; Gaps 18;

Qy      165 LIHNVREIHPNCPG-LASLELRRLSLAADPAPLPPOQOQHALLHGAPAAAPAGL- 222
      : | | | | : : | | | | : | | | | : | | | | :
Db      81 LSHLATDVH-YNPSELYSWLDNML-----SELNPPPLP-----ASSNGLD 120
      : | | | | : : | | | | : | | | | : | | | | :
Qy      223 -TLPP-----PLPDKRREHPPPCQOQOQEEHPAPQSPKAPTAETAA- 268
      : | | | | : : | | | | : | | | | : | | | | :
Db      121 PVLSPETICGPASDYDLEVPGNALYQFPALDISSSSNNQKRLKSCSPDSMTSTST 180
      : | | | | : : | | | | : | | | | : | | | | :
Qy      269 -----AAAAQAAAAAAKKEKEQKQKQDEEGHLITLLLOCAEAVNADNLDA 318
      : | | | | : : | | | | : | | | | : | | | | :
Db      181 GTQIGVIGTIVTTTTTTTAAESTRSVILLVDSQENGVRVLAHLMACAIAIOQNNLTJA 240
      : | | | | : : | | | | : | | | | : | | | | :
Qy      319 HQTLEIAELATPFGSTORVAAYFAEAMSAKLVSCLGLVAPLPQSPAAARLHGVA 378
      : | | | | : : | | | | : | | | | : | | | | :
Db      241 EALVQKQIGCLAVSQAGAMRKVATVYFAELAR-----YRLSPQN-----QIDHCLSD 289
      : | | | | : : | | | | : | | | | : | | | | :
Qy      379 AFQV-FNGISPFKFSHTANAOIAEAFEREERVHIIDIMOGLOWPGLPHILLASRPG 437
      : | | | | : : | | | | : | | | | : | | | | :
Db      290 TLOMHFEYTCRYLKFAPHTANQALILEAFEGKKRVHVIDFSMNOGLQWALMALALREG 349
      : | | | | : : | | | | : | | | | : | | | | :
Qy      438 PPRVRLTGLGA---SMEALATGKRLSDPADTLGLPFEFCA-VAEKAGVNDPEKLG- 490
      : | | | | : : | | | | : | | | | : | | | | :
Db      350 PPTFRITGIGPPAPDNDSHLHEVGCKLAQLAELAHVEEYRGFAVNSLADIASMLRLP 409
      : | | | | : : | | | | : | | | | : | | | | :
Qy      491 TRREAVAVHMLH--HSLVDYTGSDSNTLMLQRLAPKVVMVWEODLSHSGS-FLARFEVA 547
      : | | | | : : | | | | : | | | | : | | | | :
Db      410 SPTFAVAVNSVFEHLKLRGPGIEKVLGVYKQIKPVLFYVEQESNHNCPVFLDREFTES 469
      : | | | | : : | | | | : | | | | : | | | | :
Qy      548 IHYSALFDSLDASYGEDSPERHVVQOLLSEIRNVLAAGSPART-GDVYFGSGWREKLA 606
      : | | | | : : | | | | : | | | | : | | | | :
Db      470 LHYSTLFDSLE---GVPSQDKVMSEVYLGKQICNLVACGPGPRVERHETLTSQWGNRFG 526
      : | | | | : : | | | | : | | | | : | | | | :
Qy      607 OSGFRAASLAGSAAQAASLLGMFPPS-DGYTLVEENGALKGMKDLCLLTASAMR 660
      : | | | | : : | | | | : | | | | : | | | | :

```

Tue Apr 23 11:12:24 2002

Db 527 SSGIAPAHIGSNAPKQASMLSVENSQGGYRVEESNGCLMLGWHTRPLITTSAMK 581

Search completed: April 22, 2002, 19:42:46
Job time: 571 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2002, 14:31:59 ; Search time 4888.1 Seconds

(without alignments)
11846.137 Million cell updates/sec

Title: US-09-265-585c-95

Perfect score: 3510
Sequence: 1 ccgctagctcagcctactca.....tagccatcaagcaagtatla 3510

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ha:*
2: gb_hhg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_vtgo_hum:*
31: em_vtgo_inv:*
32: em_vtgo_rod:*
33: em_vtgo_hum:*
34: em_vtgo_inv:*
35: em_vtgo_rod:*
36: em_vtgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3100	88.3	3100	8	AF263457	AF263457 Zea mays
2	360	10.3	100285	8	ATF24R22	ATF24R22 Arabidopsis
3	358.4	10.2	2163	8	ATU62798	ATU62798 Arabidopsis
4	350.2	10.0	137775	2	AP003866	AP003866 Oryza sat
5	287.8	8.2	3277	8	AB048713	AB048713 Pisum sat
6	241.6	6.9	23812	8	AP002838	AP002838 Oryza sat
7	241.6	6.9	176979	8	AP001168	AP001168 Oryza sat
8	209.4	6.0	174142	2	AP003259	AP003259 Oryza sat
9	174	5.0	210827	2	AC090654	AC090654 Mus muscu
10	172.4	4.9	2709	6	AX005794	AX005794 Sequence
11	171.8	4.9	1768	6	AX005804	AX005804 Sequence
12	171.8	4.9	1872	6	TA242531	TA242531 Trilicium
13	171.8	4.9	2125	6	AX005805	AX005805 Sequence
14	170.8	4.9	56870	2	Cey39B6_3	Continuation (4 of
15	169.4	4.8	83689	8	AB017067	AB017067 Arabidops
16	169	4.8	1746	6	AX005848	AX005848 Sequence
17	155.8	4.4	800	8	AF377621	AF377621 Zea mays
18	155.8	4.4	800	8	AF377624	AF377624 Zea mays
19	155.8	4.4	800	8	AF377631	AF377631 Zea mays
20	155.8	4.4	800	8	AF377635	AF377635 Zea mays
21	155.8	4.4	800	8	AF377645	AF377645 Zea mays
22	155.8	4.4	1890	8	ZMA242530	AJ242530 Zea mays
23	155.8	4.4	2255	6	AX005806	AX005806 Sequence
24	154.2	4.4	800	8	AF377626	AF377626 Zea mays
25	154.2	4.4	800	8	AF377627	AF377627 Zea mays
26	154.2	4.4	800	8	AF377634	AF377634 Zea mays
27	154.2	4.4	800	8	AF377636	AF377636 Zea mays
28	154.2	4.4	800	8	AF377639	AF377639 Zea mays
29	154.2	4.4	800	8	AF377640	AF377640 Zea mays
30	154.2	4.4	800	8	AF377642	AF377642 Zea mays
31	154.2	4.4	800	8	AF377643	AF377643 Zea mays
32	154.2	4.4	800	8	AF377646	AF377646 Tripsacum
33	153.6	4.4	800	8	AF377633	AF377633 Zea mays
34	152.6	4.3	800	8	AF377622	AF377622 Zea mays
35	152.6	4.3	800	8	AF377623	AF377623 Zea mays
36	152.6	4.3	800	8	AF377625	AF377625 Zea mays
37	152.6	4.3	800	8	AF377628	AF377628 Zea mays
38	152.6	4.3	800	8	AF377629	AF377629 Zea mays
39	152.6	4.3	800	8	AF377630	AF377630 Zea mays
40	152.6	4.3	800	8	AF377632	AF377632 Zea mays
41	152.6	4.3	800	8	AF377637	AF377637 Zea mays
42	152.6	4.3	800	8	AF377638	AF377638 Zea mays
43	152.6	4.3	800	8	AF377641	AF377641 Zea mays
44	152.6	4.3	800	8	AF377644	AF377644 Zea mays
45	152.6	4.3	298166	2	AC087563	AC087563 Homo sapi

ALIGNMENTS

RESULT 1
AF263457 3100 bp DNA 16-SEP-2000
LOCUS Zea mays SCARECROW (SCR) gene, complete cds.
DEFINITION AF263457
ACCESSION AF263457
VERSION AF263457.1 GI:10178636
KEYWORDS
SOURCE
ORGANISM

Zea mays.
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 3100)
Lim,J., Helariutta,Y., Specht,C.D., Jung,J., Sims,L., Bruce,W.B.,
Diehm,S. and Bentley,P.N.

REFERENCE
AUTHORS
TITLE
Molecular analysis of the SCARECROW gene in maize reveals a common
basis for radial patterning in diverse meristems.
JOURNAL
MEDLINE
PUBMED
Plant Cell 12 (6), 1307-1318 (2000)
20407145
10948251

REFERENCE	2 (bases 1 to 3100)
AUTHORS	LIm,J., Helariutta,Y., Specht,C.D., Jung,J., Sims,L., Bruce,W.B., Diehn,S., and Benfey,P.N.
TITLE	Direct Submission
JOURNAL	Submitted (03-MAY-2000) Biology, New York University, 100 Washington Square East, New York, NY 10003, USA
FEATURES	location/Qualifiers
source	1..3100
organism="Zea mays"	
/db_xref="taxon:4577"	
join(<1..1562,2410..>2854)	
/gene="SCR"	
/product="SCARECROW"	
<1..>2854	
/gene="SCR"	
join(1..1562,2410..2854)	
/gene="SCR"	
/function="transcription factor"	
/note="zmscr"	
/codon_start=1	
/product="SCARECROW"	
/protein_id="AAG13663.1"	
/db_xref="GI:10178637"	
/translation="MPPPPPPPLTPYRCRCRPPPHLPPEPPSSNPHLTHYLQHDHQ EAAAMVRRKRASDMDLPPRRHYMGDSLDVYRAAAGVGSGAGSAAOCPALPT OLHOLPFAFOHABEDVYVAPHPAPAHARQOGERTASTTAWDGIITDITGSSGAAV STIOLHNVEIILHPCNPGLASLLELRSLADAPLPAPPOQOHALHGAAPAA PAGITLPPPLDPKRNHEHPPCOQDODEPHPADQSPKAPTAETRAAAAQAQA AAAKERKEEQQRKRDDEGLHLITLLQCAEAVNADNDADQTLLEIAETPEST STOVAVAFAPAMASARVSSGLGYAPLPGPSAPAAHLHGRVAAPFVNGISPEVF SHFANAIOAEPRERERVHITDIDIMOGIOMGLFHIILASRQGPFRVATLTGASM EALFATKRLSDRADTLGLPFRCAVAEKGNVDEPEKLVYTRERANAAMVHLSLTV TGSQNTLMLLIQRLAPKVVTVMEQDLSHSSFLAREVIAIHYSALFDSLDAAYGDS PERVIVAEQOLLSREIRNVLAIVGGPARVDYKFGFSWRKELAQSGFRASASLGSAAQAS LLIGMPSPDYTLVEENGALTKLWKMDICLLTASAMPRIQVPCR"	
BASE COUNT	585 a 1064 c 865 g 586 t
ORIGIN	
Query Match	88.3%; Score 3100; DB 8; Length 3100;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 3100; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
0Y	293 atgccacgcacacgcgcgcgcctctctcaatcttatctgcgcgcgcctccccc
Db	1 ATGCACACGCCACGCCCTCCCTCTCTCTCACTCTTATGCGCGCTGCCCTCCCA 60
0Y	353 caccctccctcgccctcctctcttcccaaacactctccctccactactcatgaag 412
Db	61 CACCTCCTCGCCCTCTCTCTCTCCCAACCACTCTCTCACTACTACTCCATAG 120
0Y	413 ctagaacacaaagaagccgcgcgcgcctctctctctctctctctctctctctct 472
Db	121 CTAGACACCAAGAAGCGCGCGCGCGCCGCAATGATGTCGCAAGCGCGCTCCGACATG 180
0Y	473 gaactccgcgcgcgcgcgcacagtcacagggcgaaactctctcgacgctcagggcgcgct 532
Db	181 GACCTCCGCGCGCGCGCGCGCACGTCAAGGGGCACTCTCTCGACGTACAGGGGCGCT 240
0Y	533 gccgcgcgctgt 592
Db	241 GCCGCGGT 300
0Y	593 cccacccgcgcacacacgctccccgcgcgttccagacacacgcgcgcgaggtgagctg 652
Db	301 CCCACCCAGCTCCACACTGCTCCCCCGCTTCCAGACACACGCGCGGAGGTGAGCTG 360
0Y	653 ccgcgcacccgcgcgcgcgcgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 712
Db	361 CCGCGGACCCG 420
0Y	713 acgcgctgtgtgtgacgacgacatccgcgcacatccgcgcgcgcgcgcgcgcgcgcgc 772

[illegible]

QY 1853 aggttagaagaagtaacacatctctcgatccctgactccttgctacacatgtlcaaatcttg 1912
 |||||
 Db 1561 AGGTAGGAGAGGTACACATCTCTCGATCCTGACTCTCTTGGTACCATGTCAATCTTG 1620
 QY 1913 atgcaatcattgccaacttttaagctactaaactttggttagcgaatgagacatcagt 1972
 |||||
 Db 1621 ATGCAATCATGCGCACTTTCTAGCTACTAACACTTGTAGTTCGCAATGGCACTCCAGT 1680
 QY 1973 acaactaatcttaaaaaaataactctcagaggttccctcagtaaaaaacccggttttcgga 2032
 |||||
 Db 1681 ACAACTATCTAAAAAATAATCTTCAGAGGTTCTCAGTAAAAAACCCGCTTTTGGGA 1740
 QY 2033 gctcaaaaagctgtlcatltaagacaaaccaacttcttaaggcttaaaaaaggtltaacttg 2092
 |||||
 Db 1741 GCTAAAAAGGCTCTCATTTATGACCAACCACTTTCTAGCTTTAAAAAGGTGATCTTG 1800
 QY 2093 gcaatgcttttgagagatgctgtacagaaagtactgtgtaagagagatctctcatagcc 2152
 |||||
 Db 1801 GCAATGCTTTTGACACAGTGTGTAAGTACTGTGAGAGAGATATCTCCATGCGC 1860
 QY 2153 ttgttgatcccaagaacacaaaagatagatattcgcctcgacttggttagtgagagtg 2212
 |||||
 Db 1861 TTTGTTGATCCAGACACACAAAAGATAGTATTCTGCTCGATTTGGTTAGTGGAGGTGT 1920
 QY 2213 tctgatacacttggagagatgagctgaaagttccatcatcaatgacaaacttccatg 2272
 |||||
 Db 1921 TCTATATATCATCTTGAGAGATGAGCTGAAAGTCTCATCATCATGACCAACTTTCCATG 1980
 QY 2273 gcaaaaagtttctagttccaaagtgcagagacgagatgactgagtgagcttgaatgaaata 2332
 |||||
 Db 1981 GCAAAAAGTTTCTAGTTCCAAAGTGGCAGAGCATGATTACTGAGTGCCTGAATGAGATTAA 2040
 QY 2333 ctgtcatcttctacactaaacatcatltaataataaatcatcatccgagacctaa 2392
 |||||
 Db 2041 CTGTCACTCTTCTACCACTAACCATCATTTATTATACATTAATCATCATCCGGAGCTTA 2100
 QY 2393 actaagaagaagctataataaagtgcaatcttctcaaatggtcgtccatgtgcaagtgtga 2452
 |||||
 Db 2101 ACTAGAAAGGCTAATCAAAAGTGCAATCTTTCCAATGCGCCATATGACAGTGTGA 2160
 QY 2453 catgctggccatgtacttcttgcgtgaaaccatctcgtctcaagacatgagatgaaagcc 2512
 |||||
 Db 2161 CATGCTGGCCATGTACTTTTGGTGGAACATCTCTCTCAAGCATGAGATGAAAGGCC 2220
 QY 2513 tgaactgcaatgcttctgatttgatgcaacacatattagaagaacaagctlaagcgatgcg 2572
 |||||
 Db 2221 TGAATGCAATGTCTTGATTGATGCAACCATATTATGAAAGAACGCTAAGGATGCCG 2280
 QY 2573 gtccctggcaagggaatgcatatctgcacacagacagaggaatccggaatcgatgctag 2632
 |||||
 Db 2281 GTCTGGCAAGGGCAATGCAATATGTCACAGACAGGAGATTTCGGAAATGGCTAG 2340
 QY 2633 ctggtgcaaatcgcaacggggaatataaactacatattggtcatgtattccatccccaca 2692
 |||||
 Db 2341 CTGGTGAACAATGCGACAGGGGATTAAATAACTACATTGGTCATTGATTCATCTCCACACA 2400
 QY 2693 caactcgagctggcccccaagtggtgacaatggtgagacaggaacctggaacctcgag 2752
 |||||
 Db 2401 CACTGTGAGGCTGGCCCCCAAGGTGTGACAAATGTGTGACAGAGACTGAGCACTCGGG 2460
 QY 2753 ctctctctggcgagcttggtgagggccatcaactactactgagcgctgtgtgaactcgct 2812
 |||||
 Db 2461 CTCTCTTCTGGCGGCTTCTGTGGAGGCAATCACTACTCTGGCGCTGTTCAGCTGCT 2520
 QY 2813 ggaacgagctlaagcgagagacagccccagcgacagctgctgtgagacagagctgtctgc 2872
 |||||
 Db 2521 GGAAGCGAGCTAGCGGAGAGACAGCCCGAGCGGAGCTGTGTGAGAGAGAGCTGTCTGTC 2580
 QY 2873 gcgggagatccgcaagctgtctggtcggtggtggcgccgagccgacagcggaagctcaagt 2932
 |||||
 Db 2581 GCGGGAATCCGCAACGTTGCTGGCCGTGGCGGGCGGCCCGCACCGGCGACGTCAGATT 2640

QY 2933 cgcgcagctggcgcgagaaagctggcgcaagctcgggttccgcgcgcgcctgcctccgcgag 2992
 |||||
 Db 2641 CGGCACCTGGCGCGGAGAAAGCTGGCGGAGTCCGGGTTCCCGCGCGCTGCTGCGCGGAG 2700
 QY 2993 cgcgcgagcgagcgctccctgctgctgcgcaatggttccctccgaagggatagacgtgtg 3052
 |||||
 Db 2701 CGCGCGGGCGAGCGCTCCCTGCTGCTGCGCATGTGTTCCCTCCGACGGGTACACCTGTG 2760
 QY 3053 ggaaggaacgagcgagctaaagctcgagtggaaggaacccctcgtccacacgagcgctgc 3112
 |||||
 Db 2761 GGAAGGAACGGCGGCGCTTAAGCTCGGCTGGAAGAACCTCTGCTGCTACCGCTCGGC 2820
 QY 3113 ctggcgcccatccaggtgcgcgcggtgcgttgatgagacctctgcgtcctcgtctgc 3172
 |||||
 Db 2821 CTGGCGCCCATCCAGAGTCCCGCGCGCTTGATGAGACCTCTGCTGCTCTCTGTC 2880
 QY 3173 gttgagaagcgcgacactcactgttttgcatctgtagtctgtgtgtgtgtgtgtgtgt 3232
 |||||
 Db 2881 GTTGAAGGCGCGCATCCACTGTTGTTGTCATCTGTAAGTCTGCTGCTGCTGCTGCT 2940
 QY 3233 gggagaataagaaagcggaacgtaactaattgctcgtgagtagatccatccatcagct 3292
 |||||
 Db 2941 GGGAGATTAAGAAAGCGGAACGTAATGCTCTGAGATGATCCATTCATTCAGT 3000
 QY 3293 gatagttactgtagtactaagcttaattagttcaatgtagatcgtctgttctgtgaagtg 3352
 |||||
 Db 3001 GATAGTTACTGATGTACTAAGCTTTAATTAGTTCAATGCTAGATGCTGTTCTTCAAGTG 3060
 QY 3353 tcatcgatcctgctgtcctgtcctgtccttcttctgt 3392
 |||||
 Db 3061 TCGATGCGTATCTTGTCTTGTCTGCTCTTCTTTCATTGTTG 3100

RESULT 2
 ATE24B22
 LOCUS ATE24B22 100285 bp DNA PLN 28-JAN-2000
 DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone F24B22.
 ACCESSION AL132957
 VERSION AL132957.1 GI:6434216
 KEYWORDS
 SOURCE
 ORGANISM
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (bases 1 to 100285)
 Bloeker, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetlier, F. and
 Salanoubat, M.
 JOURNAL
 Unpublished
 2 (bases 1 to 100285)
 REFERENCE
 EU Arabidopsis sequencing, project.
 AUTHORS
 Direct Submission
 TITLE
 Submitted (28-JAN-2000) MIPS, at the Max-Planck-Institut fuer
 JOURNAL
 Biochemie, Am Klopferspitze 18a, D-8152 Martinsried, FRG. E-mail:
 lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, project
 Coordinator: Marcel Salanoubat and Francis Quetlier, Gropement
 d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue
 Gaston Cremieux, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr
 COMMENT
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
 FEATURES
 Location/Qualifiers
 1..100285
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 1..524
 misc-feature
 /note="Various EST matches (T44303, H36397, N97294)
 without corresponding genemodel"
 2053..3629
 gene
 /gene="F24B22.10"

CDS
join(2053. .2499,2625. .2852,2948. .3454,3558. .3629)
/gene="F24B22.10"
/note="Contains Fructose-1-6-bisphosphatase active site
AA358-370; Prokaryotic membrane lipoprotein lipid
attachment site AA101-111"
/codon_start=1
/product="fructose-bisphosphatase precursor"
/protein_id="CAB70979.1"
/db_xref="GI:6822051"
/translation="MAATTAATTTSSHLLSSRRHVASSSQPTLSLSPRSJLPSNNGKRAP
TGVNRHGVASGVRCAVAADAETKTARKKSGEIQTLTGMLLRQEMKGEIDAEILTI
VMSISLACKQIALSVYRAGISNLTVGVAVNIQGEDQKLDIVISNEVSNCLDSGR
TGIJASEEDVPVAVESYSGNVYVDPDGGSSSIDAAVSTGSIKGIYSPTDECIVD
DSDDISALGSEORCIYVNCOPGNLLIAGYCAWSSIVFLTLGKGVSEFTDPMYG
EFVLTQENIEPKAGRIYSPNEGNYOMMDKIKYIDDLKDPGPTGKPYASRYGSLV
GDFHRTLLYGGIYICPPDAKSKNGKLLILECAPMSTIVDAQGKSGDSGRVLDIQP
TEIHQRPVLTIGSTEVEKLEKTLA"
2053. .2499
/gene="F24B22.10"
/number=1
2500. .2624
/gene="F24B22.10"
/number=1
2625. .2852
/gene="F24B22.10"
/number=2
2853. .2947
/gene="F24B22.10"
/number=2
2948. .3454
/gene="F24B22.10"
/number=3
3455. .3557
/gene="F24B22.10"
/number=3
3558. .3629
/gene="F24B22.10"
/number=4
4017. .5928
/gene="F24B22.20"
complement(4017. .4082)
/number=1
complement(join(4017. .4082,4178. .4262,4350. .4977,
5060. .5242,5424. .5658,5755. .5928))
/note="similarity to mucin MUC3, Homo sapiens,
EMBL:AF007193"
/codon_start=1
/product="putative protein"
/protein_id="CAB70980.1"
/db_xref="GI:6822052"
/translation="MGKSSRSKSGSNLIGKEVPTQVAFIVDRYLHNDRSETRALER
SEASLISDPIRMVPLNMLTDLAMLNHYSLKKOKYSLDQDKLKDOEIRQNLQ
GMENVMNTYNASLTAPPASAPTSQOKNHSISSGSLQYNTLNGMSYLNKRNVRG
NESTPSPQSQTGCKGRKGPVATAPPSRKSRITRARGTKLKROADKAANPTSETLA
VAKASASNELIGSSSVYKCLFNKADSVPTSTICEFTPKHASSGSDKNSQKEVT
PNTNCTIVTERKPTISPLKQITSYVERSHLISFSSPVKSNLKMKNKRHHVQKINF
DDTDTETCLEAPATADLVSTSPSGSEPEVDLPMDSENLDFSELVDFDGCGSANH
CISLTPTNQPTQTVSGSSPESGDCNLESDDQPLEYTSVTDVIOISRCIDPLSNSSHS
LNTDKI"
complement(4083. .4177)
/number=1
complement(4178. .4262)
/number=2
complement(4263. .4349)
/number=2
complement(4350. .4977)
/number=3
complement(4978. .5059)
/number=3
complement(5060. .5242)
/number=4
complement(5243. .5423)
/number=4

exon
complement(5424. .5658)
/number=5
complement(5659. .5754)
/number=5
complement(5755. .5928)
/number=6
5928. .6705
/gene="F24B22.30"
complement(join(6432. .7073,7206. .7519,7744. .7911,
8003. .8470,8573. .8705))
/note="similarity to delta-lactolinsectoxin precursor,
L.mactans, EMBL:LMTDLATOX"
/codon_start=1
/product="putative protein"
/protein_id="CAB70981.1"
/db_xref="GI:6822053"
/translation="WGIIDKQKQLQOEYIVYVGEPPSSSPNEPPSSIPNVEDTSED
VTPHRSRLMTKAVLTGMKTNASTLISKECNVVEQITGNSELAHLIAVAKHDFVR
NLRLREMDPDLISLKNKDNPTLSFAALGDIETEMALINMLRDLPDISENKTWPIIH
ALYGHGEWQYLFPSKTSIKDLNDQOYLNFHTMSADIVGVADVPMLDERVLYR
KELALYPNNSKALHLARKTSALSHKSOULFOQVASSMLFPAALGVLNVLILYR
SHLDLIMTYDNNRNLTFVYALYRHEINLESLIYELGIGKDLIASYKEKOSDILLIY
ARLPNPNRQVSGGAAIHMOKELIFKAVKEIYPRSYIETNTGELANDHIFTQCHEN
LRRKEGRWKETATACMLGATLIATVRAAAITTPGGNDSDGKANTLGFENFRRL
FDIFLSDSVALFSSMSMSTVIFLSIFTSRYAEEDRYDLPKLMFGSLALFISIML
LAFTFSMLIRVEKASLSLVLSCLASLTALTFAVLYPHLMFNLRYSYISMPFLGR
KRGIIRLFV"
complement(6432. .7073)
/number=1
complement(7074. .7205)
/number=1
complement(7206. .7519)
/number=2
complement(7520. .7743)
/number=2
complement(7744. .7911)
/number=3
complement(7912. .8002)
complement(8003. .8470)
/number=4
complement(8471. .8572)
/number=4
complement(8573. .8705)
/number=5
8705. .11551
/gene="F24B22.40"
10499. .11551
/gene="F24B22.40"
10499. .11551
/note="similarity to probable lectin 2 precursor, alfalfa,
PIR:T09620"
/codon_start=1
/product="putative protein"
/protein_id="CAB70982.1"
/db_xref="GI:6822054"
/translation="WASFTLSGLLVLLSTALLSISISIEATSSFFTSCKGNASE
EESDIALYGDALVDGGSILQITDVSHGGRVLYKPIESNMNNTKEVYFGFTFS
FSISPSRGRLGFVFPVNETFDHSLFQVKEDTDPDSFTQIGDSNVAVYDQATYSERI
RNFTIANLEKTEKVLIVAMINYOAGKFELEVALSKSFESVLPMLFQDIDLSOMLRD
EDEPVAIVNSYSGNVNLHSMSELEVHSEYEHSMAPVLEBOLRBEOLRKEAAKRT
SPRMREVTYCFYMTFGSGTGLVFFAMHIIYAAFKRNRLAMVNOECGIKTKRFGYKKME
KKEVYTSNADAKQERK"
10499. .11551
/gene="F24B22.40"
/number=1
13247. .13969
/gene="F24B22.50"
/number=1
13247. .14937
/gene="F24B22.50"
join(13247. .13969,14245. .14937)
/gene="F24B22.50"
/note="similarity to fructokinase, Lycopersicon

/protein_id="AA806318.1"
 /db_xref="GI:1497987"
 /translation="MASGDPGNGPDPHPSPLRTSSGSSSSNNKRCPPPPPPPLVWY
 RKRLASMSNNPDYNNSSRPFRKRSGLDDSNINVTQOPLSLAATVSSQNPPLS
 VCGSGGLPFPSDGRGGNNMWSVQPMQDSSSSASPTAVWD1IKRLHSTVSVIP
 OLIONVNDIIPPCPNLGALEYLRSLMILDPSSSDSPQTEEPYLYQ1SNNSPPO
 QOQOQOQOQOQOQHPKPPPIQOOERENSTADPAQPEYTVATPVAOVN1EALRERE
 E1IKQKQDEBGLITLILLOCAEAVSADNLEANKLILETSQISLPTGTAQVAAVE
 SEAMASRLNCSGLGITVALPDRMNPQHSKLNKAVFTGIVSPVKE1SFTNNAIO
 EAEFKEDSVH1IIDIDMOGLQWPGILFVILILSRGQPHVHVL1GTGSMELQATGRL
 SDPFDLGLPFEPICPQPAERKVGNDLTERLNRKREAVAVHMLQIXVDVSDHATLML
 LQRLAPRVVTVVEDDLSHAGSPFEGRAFEAL1HYXSLSDSIS4YAGSESPERRHVEOL
 LSKRIKRVVLAVGSDPGRSGEVKFPSPMRKQCCFGKISLGNNA1QATQTLILGNFPSDG
 YTVLDNGTLLKLGKMDLSLITASAMTPRS"
 1581..1695
 /gene="SCARECROW1"
 exon .125
 1696..2125
 /gene="SCARECROW1"
 571 a 546 c 439 g 607 t

Query Match	10.2%	Score 358.4;	DB 8;	Length 2163;
Best Local Similarity	59.3%	Pred. No. 1e-33;		
Matches 696; Conservative	0;	Mismatches 451;	Indels 27;	Gaps 4

[illegible]

QY	1422	ccgcgcaggttcgcagggtgtgttcaacgcgtacacgcccttcgtcaagttcttcgcacttcacccg	1481
Db	1151	CT-----GGGTTTCAGGCTCTTTAAATGGGATAAGCCCTTTACTGAAATTCCTACACTTTACAG	1207
QY	1482	ccaacacagccatccatccagaaagcgcttcgaagcggaagaaagcgctgcacatccatccagccctg	1541
Db	1208	CGAATCAGGGCGATTCAGGAAGCATTTTGAGAAAGAAACAGTGTACACATCATTTGACTTGG	1267
QY	1542	acatcatgcaagggagctgcagttgacgcggggagctcttcaatctctgcttcacgcgcccgagg	1601
Db	1268	ACATCATGCAAGGAGGACTTCATTCGCTGCGGTATTAATTCACATTTCTTCTCTAGACCTCGAG	1327
QY	1602	gcccgcgcaggaagtgaagctacccgcctcggggcgctgcattgaagcgcttcgaagcgcacgg	1661
Db	1328	GACTCTCCACAGCTCGACACTCAGCGGACTTGCTGTAATTCATGGAACCTCTTCAGGCTACAG	1387
QY	1662	ggaagcgctctccgaatlctgcgcgacacgcgctcgcccttcgattctgcgcgcctg	1721
Db	1388	GGAACCGCTTTTGGATTTTCACAGATTAACCTTGGCCTCGCTTTAGTTTGCCCTTTAG	1447
QY	1722	ccgaaagagcgccgaatgttgaccgcggaagaagctaggggtccagggcggaagcgcgctg	1781
Db	1448	CTGGAAGAGTTGGAAACTTGGAACTGGAAGATCAATGTGGAAGAAAGGAAGCTGTGG	1507
QY	1782	ccgtccacctgctgcagccactgcgtcttaagcagctcaactgctgcgcacccatccacacgcgtct	1841
Db	1508	CTGTTCACCTGGCTTCAACATTCCTTTATGATGTCACTCGCTCATGTACACACACTCTCT	1567
QY	1842	ggctcatccaaaagttagaaagagttcacaccatct	1875
Db	1568	GGTTACTCTCAAGCTAAATAAATAACATTCACCTTTT	1601

RESULT	4
LOCUS	AP003866
DEFINITION	AP003866 137775 bp DNA HTG 10-JUL-2001
ACCESSION	Oryza sativa chromosome 7 clone OJ1092.A07, *** SEQUENCING IN
VERSION	PROGRESS ***, in ordered pieces.
KEYWORDS	AP003866
SOURCE	AP003866.1 GI:14646799
ORGANISM	HTG; HTGS_PHASE2.
	Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1092_A07.
	Oryza sativa
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
	Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE	1 (bases 1 to 137775)
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE	Oryza sativa nipponbare(GAS) genomic DNA, chromosome 7, BAC
	clone:OJ1092_A07
	Published only in Database (2001) In press
JOURNAL	2 (bases 1 to 137775).
REFERENCE	Sasaki,T., Matsumoto,T. and Yamamoto,K.
AUTHORS	Direct Submission
TITLE	Submitted (09-Jul-2001) Takuji Sasaki, National Institute of
JOURNAL	Agrobiological Resources, Rice Genome Research Program, Kannondai
	2-1-2, Tsukuba, Ibaraki 305-8602, Japan
	(E-mail:tsasaki@agr.affrc.go.jp, URL:http://jgp.dna.affrc.go.jp/,
	Tel:81-298-38-7441, Fax:81-298-38-7468)
	The nucleotide sequence of this BAC clone was generated by
	combining Monsanto and RGP-Japan sequencing data.
	NOTE: It currently consists of 1 contigs. Gaps between the contigs
	are represented as runs of N. The order of the pieces is believed
	to be correct as given, however the sizes of the gaps between them
	are based on estimates that have provided by the submitter. This
	sequence will be replaced by the finished sequence as soon as it is
	available and the accession number will be preserved.
	* NOTE: This is a 'working draft' sequence.
	* This sequence will be replaced
	* by the finished sequence as soon as it is available and
	* the accession number will be preserved.
FEATURES	Location/Qualifiers

[illegible]

KEYWORDS	pisum sativum (cultivar:Alaska) 5 days old cDNA to mRNA, clone_lip:lambda GEX5 pea shoot cDNA clone:pSCRI and pSCR2.		
SOURCE	pisum sativum		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; pisum.		
REFERENCE	1 (bases 1 to 3277) Sassa,N., Matsushita,Y., Nakamura,T. and Nunoya,H. The Molecular Characterization and In situ Expression Pattern of pea SCARECROW Gene Plant Cell Physiol. 42 (4), 385-394 (2001) 2 (bases 1 to 3277) Nunoya,H., Matsushita,Y. and Sassa,N. Direct Submission Submitted (08-SEP-2000) Hiroshi Nunoya, Tokyo University of Agriculture and Technology, Gene Research Center; 3-5-8 Saiwai-cho, Fuchu, Tokyo 183-8509, Japan (E-mail:nunoya@cc.tuat.ac.jp, Tel:81-42-367-5563, Fax:81-42-367-5563)		
FEATURES	Location/Qualifiers 1..3277 /organism="Pisum sativum" /cultivar="Alaska" /db_xref="taxon:3888" /clone="pSCRI and pSCR2" /clone_lip="lambda GEX5 pea shoot cDNA" /dex_stage="5 days old" /note="PCR-paSCR41, PCR-paSCR33" 322..2781 /gene="psscr" /gene="psscr" 322..2781 /protein_id="BAB39155.1" /db_xref="GI:13365610" /translation="MAACALFNGVGGNTPDETNNNSNSSNSTEDFHNPQOP HSEERILRRMASEMELQHNNNNNNDYHRESRTNNSTSLNSCLPATQKGVTTT TTTLASSGNNNNNNNNNNHYHNNNNNSIINNNNNVALSROVAIONPEVYVTTN YSTMILPSSCSNNLNNSSTSAANTHYQOLPEONOTPLTCGSGILPSPSONOTN RTNNSSNNNNNTYVDYVSSPSMEETATITWIDGILKDLHTSNSTISPOLINN VRETLTPCNPNLALVEHRILRLTEPTQPERKRNSTEGSVNVCNVLASVNNSS SVKLNRNDVDVPTSLFSDSSTLNONOMNPNMGAIINNNDPSVLTLPQO PLETOOQOQHOLOQDEPLDPAFTTTTSELALARKKEIKEOKKDEGJLLTL LLOCAEVASEINLEOAKMLLEISOLTPEGTSORVAAYFSAISRLVSSCGIYAA TLPVSSHTPNORVYASAFVYENGISPPVKSHPHTANAOIOEAFERERVHIIDIDMO GLQMPGKPHILASRPGRPYVRLTGLCTSMETLEATGRKLSDPANKIGLPEPPVAB KVGIDIEKLNTSKSEAVNVAHYMHDSITDVTGSDNTLMLLQRLAPRVVTEODSS AGSFLGFGEVAIHYYSALFSLGSSSYSESEERHVVQQLSRIRNVLAVGGSRSG EIFEHMRKELQOCGFGRGVLASGNAATQASLLILGMFSEGYTLEDNGILTKGKDLCL LTASARPPYHNTIIPHNH"		
BASE COUNT	1038 a	665 c	562 g 1012 t
ORIGIN			
Query Match	8.2%;	Score 287.8;	DB 8; Length 3277;
Best Local Similarity	63.2%;	Pred. No. 1.8e-25;	
Matches 461; Conservative	0;	Mismatches 262;	Indels 6; Gaps 1;
QY	1133	aagagagcggaagggagcgagcgagcggaagcgagcgagggagggcctcactgctg	1192
DB	1606	AAGAGAGAAAGGAGATTAAAGAACAGAGAAAGAAAGAGGTTTGCACCTTCTTA	1665
QY	1193	acgctgctgcgcagctgcgcgcgagcgctgaacgcggaacaacctggagcgagcgaccag	1252
DB	1666	ACACTTCTTCTCCCAATGCGCAGAGCGTGTATCAGCAGAGATCTAGAAACAAGCAACAG	1725
QY	1253	acgcctgcgtgagatcgcgagctagcgagcgcttcgtgcacacttcgacgcgagcgctggcc	1312
DB	1726	ATGTTACTCGAGATTTCGACACCTCTCGACCTCCATTTCGGGAGATTCGCGACAAAGATGACA	1785

[illegible]

RESULT	6		
AP002838/c			
LOCUS	AP002838	22812 bp	DNA
DEFINITION	Oryza sativa genomic DNA, chromosome 6, BAC clone:OSJNbA0038F22, complete sequence.		

VERSION	AF002838.1	GI:9711845
KEYWORDS	HTG.	
SOURCE	<i>Oryza sativa</i> (cultivar:Nipponbare) DNA, clone:OSNba0038F2.	
ORGANISM	<i>Oryza sativa</i>	

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE
1 (sites)	Sasaki, T., Matsumoto, T. and Yamamoto, K.	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, BAC clone:OSJNBa0038F22	Published Only in Database (2000) In press	2 (bases 1 to 23612)

AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2000) Takuji Sasaki, National Institute of

Agrobiological Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tssasaki@affrc.go.jp, URL: <http://rpg.dna.affrc.go.jp/>)
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
The orientation of the sequence is from M13rev to -21M13 of the
clone.

```

FEATURES
SOURCE
Location/Qualifiers
1. 23812
/organism="Oryza sativa"
/cultivar="Nipponbare"

```

```

/dl_xref="taxon:4530"
/chromosome="6"
/clone="OSJNBa0038F22"
BASE COUNT      6315 a      5732 c      5372 g      6393 t
ORIGIN

```

Query Match	6.9%	Score 241.6;	DB 8;	Length 23812;
Best Local Similarity	57.6%	Pred. No. 2.1e-20;		
Matches 515; Conservative	0;	Mismatches 364;	Indels 15;	Gaps 4;

961 gctcctccccgcacgcgttcgcgacaaagcgcgcacgagatccaccgcgtgcca 1020

Db 7867 GGTGCCAGCGCGCGGCGTTCGCCGACGGCATGGCCGCCCGCGGGGCCATCGCTGGCCCC 7808

QY 1021 gcagcaacagcaggaacccgcatccgcgcgcgcagtcgcccaagcccgaccgcgca 1080

Db 7807 GGCCTAGAGAGAGATTACAGCAATGCCGTCGACCTGTCCTCTGTCGGGACACGA 7748

Qy 1081 aaagaccgcagcgcgcgcgcgcacaaagacgctgtctgcgcgcgccaaagagcg 1140

Db 7747 CGCGGGCTGTCGCC-GATGATGCCCGCGCCGAGCCGC-----CCGAGTTCGACGGCG /694

[illegible][illegible]

Db 7533 CACCGCGTGGCCGACCTCCCTTCGCGCGGCAACACGAGGCGGCCACTACTACCTGGC 7574

QY 1261 gaagatcgcgcgaactagcgaacgcgttcg----cacttcgacgcagcgcgtgcccgccta 1317

Db 7573 CCGGCTCGGGGAGATGGCTTCGGCCGGGGGGCCACGCGGATGCACCGGTTGGCCGCTTA 7514

QY 1318 cttcgcgagaccatgtctgcgcgcgtctcagctcctctgcctgtgctgt-----acgc 1371

Db 7513 CTTTACCGAGGGCGCTCGCGTCGCGCATGTGGCCGACATGTTGCACATCGG 7454

QY 1372 gccgcgtccgcccggcgtcccccgcgcgcgtctcacagcgcgatcgccgcgcgtt 1431
||||| |||| | ||| |||| | ||| |||| | ||| |||| |

Db 7453 CCCGCGGGGAGCTCACCAGACAGCCCTTCGGCGGGCGGCGAGACAGAGCCATGGCGCT 7394

QY 1432 ccaggtgcttaacgcatacggcccttcgtcaagtcttcgcacttaccgcgaacacaggc 1491

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

7200 cccccctttatcccattatcccccccccccccccccccccccccccccccccccccc 7234

1402 catcgaagagacgttcgaacgagagagagagagatgacacatcatcgaacgtcacatca 1551
 393 GCGGAACTCAACGCCCAACGCGCCATCCCGAGGTTCCAGCACTCAAGCCTCAACGAG 7534

Dh 7333 CTTCTCCGAGTTCGAGGGGCACGAGCGCCTCACGTCATGTGACATGAAGCA 7274

1552 qqqqcbcaqlvqqcqqgqcbtcttccacatccttgcctccqccccqgggqccqccag 1611

Db 7213 GGGGCTCCATGCGCGGCGCTTGCTCCAGAGCCTGGCGCGCGGCGGCGGTGCTCCGCGCGCA 7214

QY 1612 ggttaggcacccggcctcgggagctccatgtaggcgtcgaagcacaagggaagcgct 1671

Db 7213 CGTGGCGATCACCGAGTTCGGCGAGTCGAGCGAGGACCTGCAGAGACGGGCGCGGCT 7154

QY 1672 ctccgatttcgcgcgaacagctcgcgcctcgcccttcgafttctgtgcgccgtcgcgcgaagaagc 1731

Db 7153 GGGGCGGCTCGCCGCGCGGCGCTGGCGTTGAGTTCACAGCCGCGGTGACCGGCT 7094

QY 1732 cggcaatgtttgacccgcgagaaactagsgtctcagaagcgaggccgtccgtccaattg 1791
|| || || || || || || || || || || || || || || || || || || || || || || ||

Db 7093 CGAGACGTCGCCCTGTGGATGCTCCACGTCAAGCGCGGAGATGCGTGCGCCGTGACTG 7034

QY 1792 gctgcaccactgcgtctacgaagctacgtgctccgactccacaacagctctcggct 1845

DB / 033 66TCCCTCCCAAGACACCGCCCTGCTCCCGAGACACCCCGCCGACGACTTCCT 0360

RESULT 7

AP001168/c	LOCUS	AP001168 176979 bp DNA PLN 16-FEB-2000
	DEFINITION	Oryza sativa genomic DNA, chromosome 6, clone:P0425F02.
	ACCESSION	AP001168
	VERSION	AP001168.1 GI:6983854
	KEYWORDS	
	SOURCE	
	ORGANISM	Oryza sativa (cultivar:Nipponbare) DNA, clone:P0425F02.
		Oryza sativa
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartioideae; Oryzaceae; Oryza.
	REFERENCE	1 (bases 1 to 176979)
	AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.
	TITLE	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC clone:P0425F02
	JOURNAL	Published Only in Database (2000) in press
	REFERENCE	2 (bases 1 to 176979)
	AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.
	TITLE	Direct Submission
	JOURNAL	Submitted (14-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@agr.affrc.go.jp, URL:http://www.dna.affrc.go.jp:82/, Tel:81-298-38-7441, Fax:81-298-38-7468)
	COMMENT	The orientation of the sequence is from SP6 to T7 of the PAC clone. Genes were predicted from the integrated results of the following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October1998 version). The genomic sequence was searched against the non-redundant database NRP(PTR,SWISSPROT, GENPEPT, PDB) from MAF DNAbank and the cDNA sequence database at RGP. Protein similarities of the coding regions were searched against NRP with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN2.0 with the corresponding DDBJ accession no. and RGP clone ID. Detailed information on assemble quality together with annotation of this entry at http://www.dna.affrc.go.jp:82/genomicdata/genomef1.nished.html.
FEATURES	Source	Location/Qualifiers
		1..176979
		/organism="Oryza sativa"
		/cultivar="Nipponbare"
		/db_xref="taxon:4530"
		/chromosome="6"
		/clone="P0425F02"
		36..254
		/note="EST AU082143(E4368) corresponds to a region of the predicted gene.: hypothetical protein"
		/codon_start=1
		/protein_id="BAA90790.1"
		/db_xref="GI:6983855"
		/translation="MAELPPNDATIVMDKEEVEEKEIVEEKPSPAPAEVNNTEVEE ETELVNTEVDTGTEVKNSEEEKPIOS"
		join(1092..1123,1482..1740,1821..2060)
		/note="hypothetical protein"
		/codon_start=1
		/protein_id="BAA90791.1"
		/db_xref="GI:6983856"
		/translation="MDDQLDRISNCAKPNYKSTDRKGSATYCGFGDGDGDAITHR GRLPRRVAELIHOOVCALVAPNSPSSRCSLSTCRRRSDTMSSEFCITGIFLIDOGI SOHAGROORRDSASMAAIRKSTAILLRAHQSRRRRPIHOKKAVLETSOHPLTASA TASPLEGSKRISROIIT"
		complement(join(2859..3254,3286..3498,3504..3752,3923..4483,4546..4662))
		/note="hypothetical protein"
		/codon_start=1
		/protein_id="BAA90792.1"
		/db_xref="GI:6983857"
		/translation="WKSGRKAGRPPTAQAARAGRGGERPVDRAHHARSRGRTG RCTRAVGRGCGGPTKSLTARGCTRAREGIDRGRLDALVLAFTWRLRGCHGRRE YDDAGRNGRSTASDGNHGDGTGSEHTGLHYTRGDEPTAIRRLREGSGESRR OPAAGEENGDEVARGOPPARASTRLRESASVGLGGTTPSEAGDERPRPARAATAA SRRRATATEEEBEGARGVPRVLVCTGSSGTHATTNGDEBERRRRRETREKRELGCGG
		complement(join(35125..36722,36911..37108,37364..38459))
		/note="Similar to 22 kDa kafirin cluster; Ty3-Gypsy type (AP061282)"
		/codon_start=1
		/protein_id="BAA90795.1"
		/db_xref="GI:6983860"
		/translation="MVTNRSGNCPNPNNNNGENPTLAQVLAOQTOLNMMMOQFO NOONOGNNHVPPOKLAEPILRVPRPVSSTTNPEASDMHATEKIDILLOCDNDOKY SFASHOIHGPAASBMDPFRILNRTAEPITMLEPTAARKHITISGVYSLKKEFRSLT QGSRSVETLYLHEFNRLARVAPEDVRNDEEKEERFLGDELDELSTPLMGVQIPQKLY DKALROEDKTYNMEOKRRRIAQEFYQGNOKRPLTLTGPQSMQSGSSSVVRQROFV NNNAGNNIRNQAQPRVVAAPAOQOPAKREOCNRTYVPCGDPCHYADCKCPKPRVAV PAONNASVAPSKARVNVAAAEADADAVTLIGMDWLAHRGVIDCANRKYTLNSYGR VVYTHALASIESRLILMOITLEETPIRYEPEDVPPDLPALSRKCYNCACTEBROJPL ELCKEPRNLIGIYGRFVAALAEKPLILDOYVEAONNDPIDEIKKRNKCAATGFL EDEGTVWLSRGICVDPNKKDKATILKEADHTLYSHPGSTKMYODLREFFWASAKR EIAIYVAVCDVQCQKAEHQKPAGLQPLKIPMKWEIEMDITGLPTSSCHDSITW VIVYRLRVKVAHFIVKTYSGSLAEIYMARIVCLAGVCPKIVSDGSOFTSNFMKL QEEMGSKLNFSTAYHPQDQTEBERVNOILEDMPACALDLDGSMNDLNPAAEYSYNS YQALIQMAPEYALVGRKCTPILMDQDGEVQVGTGILPRAEKVYVIVIERLVAASR HKSAYADRNDLSPEDDYLYLRYTPLRGCHRHRTGKGLAPRVGVYKTASRGEVAY QLELPQSLACVHNHFVYSQKLCURVTEANLQLLEVDDITLYVKKPIRLIETNRR TRNVVIFRCYVQWSNHSSEESTWEREDELKSAPHHQLFASSESGDSV"
		complement(38728..39519)
		/note="Similar to gypsy-type retrotransposon RIRBB DNA, internal region (AB014741)"
		/codon_start=1
		/protein_id="BAA90796.1"
		LGFKWAANSWERCTDIGRSRRELALGACGONGGGGMMTPAGRKEKEGKRGILAPL LGKGGGSGATROREGG/CLEPLASASGSGAEMTAMTAMAGFGATLAAAGRSAT TARGKORARASARAREHAGARAAAREARTRTLAGGALAHAAAGCGSGRA GRGSRARARAGARERRRGERGESGAPGEARARERRKAPAGREGNGRGFGPSNG DAN"
		complement(9848..12845)
		/note="Ty3' LTR"
		complement(join(12857..16323,18384..18477,19513..20442))
		/note="Similar to 22 kDa kafirin cluster; Ty3-Gypsy type (AP061282). Internal stop codon:15317-15319; probably inactive because stop codons are included in CDS"
		/codon_start=1
		/pseudoc
		complement(join(20711..20932,23987..24562))
		/note="Similar to gypsy-type retrotransposon RIRBB DNA, internal region (AB014741)"
		/codon_start=1
		/protein_id="BAA90793.1"
		/db_xref="GI:6983858"
		/translation="MLARKPEYLPGEVGFVOELRTMSFVIGCTAPYAOIPDHRBE KCRVKTLSNSEDVPSVMPFAGGRNTIHACQVARIAGELREYSDDLADTEYRH PROPGSDRGVLEFEGIENTATRHIVEMAMDESARAFVLAODREDRNGKICK LEDVDRLEKRIALKGAPQKARITLARKKALVPRYQIAPKRVYKEVYAPAD TPVLIISDEEESKRNSKTEWVAIQDDEDEPNEPSINLCEELE"
		complement(25067..28064)
		/note="Ty3' LTR"
		join(28483..28975,29105..30435)
		/note="Similar to 22 kDa kafirin cluster; Ty3-Gypsy type (AP061282)"
		/codon_start=1
		/protein_id="BAA90794.1"
		/db_xref="GI:6983859"
		/translation="MKRRITAPVLTLPDIRKDEIFPYDASROGLGCVLMQEKRVAY ASROLRPEVYPTPDLLELAUVALKIMRHVILIGNCEVYTOHRSKLYETOTELNM RORRMELIKDYDGIHYHGKANVAVDALSHKVCNCAIOIRPDODICELEKIRLT VYQSDDOGTWCGSRICVYPAKKELDILIKESASISTHSGSTKMTDITAPFWAG MKROYAETVALCDICQKRAHQRPALDLPILPEKMEIEGMDITDILPRPSGD SIWIVTRLTRKSAHFVPVKTTFDEKLAELYMHHVRCFCPKRIYSDRLOTPSRM KOLHEALGTDLNFSTAYHPDQDTEBERVNOILEDMPACALDLDGSMNDLNPAAEYSY NSYQASIKNSPNAMFGKRCRTPLCNNEGVALVFPDILKSAEBOVKILRRKTA OSROKNYADNRDRDLFEFGKGYHLRVSPRGMRPFMSKLAIPRYGPLYLITARGE VAYOLEPDEGLADVHNHFVYSQKLCURVTEVQVPLDNLITLERLTYKRPVILEET EKOTRKTITKYKQVMSHSEDAETWEREDLRAPEPELL"
		complement(32146..35112)
		/note="Ty3' LTR"
		complement(join(35125..36722,36911..37108,37364..38459))
		/note="Similar to 22 kDa kafirin cluster; Ty3-Gypsy type (AP061282)"
		/codon_start=1
		/protein_id="BAA90795.1"
		/db_xref="GI:6983860"
		/translation="MVTNRSGNCPNPNNNNGENPTLAQVLAOQTOLNMMMOQFO NOONOGNNHVPPOKLAEPILRVPRPVSSTTNPEASDMHATEKIDILLOCDNDOKY SFASHOIHGPAASBMDPFRILNRTAEPITMLEPTAARKHITISGVYSLKKEFRSLT QGSRSVETLYLHEFNRLARVAPEDVRNDEEKEERFLGDELDELSTPLMGVQIPQKLY DKALROEDKTYNMEOKRRRIAQEFYQGNOKRPLTLTGPQSMQSGSSSVVRQROFV NNNAGNNIRNQAQPRVVAAPAOQOPAKREOCNRTYVPCGDPCHYADCKCPKPRVAV PAONNASVAPSKARVNVAAAEADADAVTLIGMDWLAHRGVIDCANRKYTLNSYGR VVYTHALASIESRLILMOITLEETPIRYEPEDVPPDLPALSRKCYNCACTEBROJPL ELCKEPRNLIGIYGRFVAALAEKPLILDOYVEAONNDPIDEIKKRNKCAATGFL EDEGTVWLSRGICVDPNKKDKATILKEADHTLYSHPGSTKMYODLREFFWASAKR EIAIYVAVCDVQCQKAEHQKPAGLQPLKIPMKWEIEMDITGLPTSSCHDSITW VIVYRLRVKVAHFIVKTYSGSLAEIYMARIVCLAGVCPKIVSDGSOFTSNFMKL QEEMGSKLNFSTAYHPQDQTEBERVNOILEDMPACALDLDGSMNDLNPAAEYSYNS YQALIQMAPEYALVGRKCTPILMDQDGEVQVGTGILPRAEKVYVIVIERLVAASR HKSAYADRNDLSPEDDYLYLRYTPLRGCHRHRTGKGLAPRVGVYKTASRGEVAY QLELPQSLACVHNHFVYSQKLCURVTEANLQLLEVDDITLYVKKPIRLIETNRR TRNVVIFRCYVQWSNHSSEESTWEREDELKSAPHHQLFASSESGDSV"
		complement(38728..39519)
		/note="Similar to gypsy-type retrotransposon RIRBB DNA, internal region (AB014741)"
		/codon_start=1
		/protein_id="BAA90796.1"

147	Y	1141	gaagagagagagcagcgcggaagcagcgagagagagcgctcaactctgctgacgctgct	1200
148	Y	1081	agagagcagcgagcgcgcgcgcgcaaacagcagctgctgtgagcgcgcaagagcg	1140
149	Db	174805	cggcgcgctgctgctgcg-gatgatgctccggcgccgacgccc-----ccgaatttcagccgacg	174752
150	Y	1081	agagagcagcgagcgcgcgcgcgcaaacagcagctgctgtgagcgcgcaagagcg	1140
151	Db	174865	gcccAAGAGGAGGAGTGCACGCAAGTCCGCGTCCACACTGCTGCTCGTCGCGACACGGA	174806
152	Y	1021	gcagcaacagcagaggaacgcgacatcgcgcgcgcagctgcgcccaagcccgcgcgcgga	1080
153	Db	174925	ggTGCCGAGGCGGGGCGGTTCCGCGACGAGTGCCTGGCGGGGCGCATGCGTCTGGCGCG	174866
154	Y	961	gtccctcccccgcacacgccttcgcgaaagcgcgccagcagcgacatccacgcgcgtgcca	1020
155	Query Match		Similarity 57.6% Pred. 8.9e-21; Score 241.6; DB 8; Length 176979;	
156	Best Local		Matches 513; Conservative 0; Mismatches 364; Indels 15; Gaps 4;	
157	CDS		note="EST D15612(C0952) corresponds to a region of the	
158	CDS		note="hypothetical protein"	
159	CDS		note="hypothetical protein"	
160	CDS		note="EST D15612(C0952) corresponds to a region of the	
161	CDS		note="EST D15612(C0952) corresponds to a region of the	
162	CDS		note="EST D15612(C0952) corresponds to a region of the	
163	CDS		note="EST D15612(C0952) corresponds to a region of the	
164	CDS		note="EST D15612(C0952) corresponds to a region of the	
165	CDS		note="EST D15612(C0952) corresponds to a region of the	
166	CDS		note="EST D15612(C0952) corresponds to a region of the	
167	CDS		note="EST D15612(C0952) corresponds to a region of the	
168	CDS		note="EST D15612(C0952) corresponds to a region of the	
169	CDS		note="EST D15612(C0952) corresponds to a region of the	
170	CDS		note="EST D15612(C0952) corresponds to a region of the	
171	CDS		note="EST D15612(C0952) corresponds to a region of the	
172	CDS		note="EST D15612(C0952) corresponds to a region of the	
173	CDS		note="EST D15612(C0952) corresponds to a region of the	
174	CDS		note="EST D15612(C0952) corresponds to a region of the	
175	CDS		note="EST D15612(C0952) corresponds to a region of the	
176	CDS		note="EST D15612(C0952) corresponds to a region of the	
177	CDS		note="EST D15612(C0952) corresponds to a region of the	
178	CDS		note="EST D15612(C0952) corresponds to a region of the	
179	CDS		note="EST D15612(C0952) corresponds to a region of the	
180	CDS		note="EST D15612(C0952) corresponds to a region of the	
181	CDS		note="EST D15612(C0952) corresponds to a region of the	
182	CDS		note="EST D15612(C0952) corresponds to a region of the	
183	CDS		note="EST D15612(C0952) corresponds to a region of the	
184	CDS		note="EST D15612(C0952) corresponds to a region of the	
185	CDS		note="EST D15612(C0952) corresponds to a region of the	
186	CDS		note="EST D15612(C0952) corresponds to a region of the	
187	CDS		note="EST D15612(C0952) corresponds to a region of the	
188	CDS		note="EST D15612(C0952) corresponds to a region of the	
189	CDS		note="EST D15612(C0952) corresponds to a region of the	
190	CDS		note="EST D15612(C0952) corresponds to a region of the	
191	CDS		note="EST D15612(C0952) corresponds to a region of the	
192	CDS		note="EST D15612(C0952) corresponds to a region of the	
193	CDS		note="EST D15612(C0952) corresponds to a region of the	
194	CDS		note="EST D15612(C0952) corresponds to a region of the	
195	CDS		note="EST D15612(C0952) corresponds to a region of the	
196	CDS		note="EST D15612(C0952) corresponds to a region of the	
197	CDS		note="EST D15612(C0952) corresponds to a region of the	
198	CDS		note="EST D15612(C0952) corresponds to a region of the	
199	CDS		note="EST D15612(C0952) corresponds to a region of the	
200	CDS		note="EST D15612(C0952) corresponds to a region of the	

[illegible]


```

Oy 68 gactgactgactgtgtgtgtgtgtgtatcagcagccgcgcgcgcgaacaacgcaa 127
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16538 cccccccccccccccccccccccccccccccccccccccccccccccccccc 16597

Oy 128 actgtccctccctccctccctccctccctccctccctccctccctccctccct 187
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16598 CTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 16657

Oy 188 cggcgagctctctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 247
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16658 cccccccccccccccccccccccccccccccccccccccccccccccccccc 16717

Oy 248 cctctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 307
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16718 cccccctcccccccccccccccccccccccccccccccccccccccccccccc 16777

Oy 308 cctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 367
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16778 cccccccccccccccccccccccccccccccccccccccccccccccccccc 16837

Oy 368 cctctctctccccaactctctctctctctctctctctctctctctctctctct 427
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16838 cccccccccccccccccccccccccccccccccccccccccccccccccccc 16897

Oy 428 gccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 487
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16898 cccccccccccccccccccccccccccccccccccccccccccccccccccc 16957

Oy 488 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 547
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16958 cccccctccccccccccccccccccccccccccccccccccccccccccccc 17017

Oy 548 ggtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 607
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17018 cccccccccccccccccccccccccccccccccccccccccccccccccccc 17077

Oy 608 cagctgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 667
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17078 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17137

Oy 668 ccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 727
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17138 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17197

Oy 728 ggcatactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 787
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17198 cccccccccccccccccccccccccccccccccccccccccccccccccccc 17257

Oy 788 atccacaagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 847
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17258 cccccccccccccccccccccccccccccccccccccccccccccccccccc 17317

Oy 848 ctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 907
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17318 cccccccccccccccccccccccccccccccccccccccccccccccccccc 17377

Oy 908 cagcagcagctctctctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 967
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17378 cccccccccccccccccccccccccccccccccccccccccccccccccccc 17437

Oy 968 cccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1027
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17438 cccccccccccccccccccccccccccccccccccccccccccccccccccc 17497

Oy 1028 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1077
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 17498 cccccctccccccccccccccccccccccccccccccccccccccccccccc 17547

```

RESULT 10
 AX005794
 LOCUS AX005794 2709 bp DNA PAT 24-AUG-2000

```

DEFINITION Sequence 3 from Patent WO9909174.
ACCESSION AX005794
VERSION AX005794.1 GI:9928799
KEYWORDS
SOURCE
  ORGANISM
    Triticum aestivum
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
    Pooideae; Triticeae; Triticum.
  1 (bases 1 to 2709)
REFERENCE
  Harberd,N.P. and Peng,J.
  Genetic control of plant growth and development
  Patent: WO 9909174-A 3 25-FEB-1999;
  NICHOLAS PAUL (GB); PENG JINRONG (GB)
  HARBERD
FEATURES
  source
    misc_feature
      6
      /note="n is any nucleotide"
      /organism="Triticum aestivum"
      /db_xref="taxon:4565"
    misc_feature
      11
      /note="n is any nucleotide"
    misc_feature
      46
      /note="n is any nucleotide"
    misc_feature
      85
      /note="n is any nucleotide"
    misc_feature
      107
      /note="n is any nucleotide"
    misc_feature
      407
      /note="n is any nucleotide"
    misc_feature
      417
      /note="n is any nucleotide"
    misc_feature
      430
      /note="n is any nucleotide"
    misc_feature
      436
      /note="n is any nucleotide"
    misc_feature
      448
      /note="n is any nucleotide"
    misc_feature
      452
      /note="n is any nucleotide"
    misc_feature
      490
      /note="n is any nucleotide"
    misc_feature
      498
      /note="n is any nucleotide"
    misc_feature
      556
      /note="n is any nucleotide"
    misc_feature
      562
      /note="n is any nucleotide"
    misc_feature
      727
      /note="n is any nucleotide"
    misc_feature
      769
      /note="n is any nucleotide"
    misc_feature
      1444
      /note="n is any nucleotide"
    misc_feature
      1450
      /note="n is any nucleotide"
    misc_feature
      1556
      /note="n is any nucleotide"
    misc_feature
      1742
      /note="n is any nucleotide"
    misc_feature
      1808
      /note="n is any nucleotide"
    misc_feature
      1900
      /note="n is any nucleotide"
    misc_feature
      1940
      /note="n is any nucleotide"
    misc_feature
      1982
      /note="n is any nucleotide"
    misc_feature
      2438
      /note="n is any nucleotide"
    misc_feature
      2604
      /note="n is any nucleotide"
    misc_feature
      2638
      /note="n is any nucleotide"

```


[illegible][illegible]

Db	917	TCCTCGACGGCGGCTTTCGACGACTCCTCCACGCGGACTTCAGAGTCT-----	967
QY	1452	ggccctctgfcgaagttctctgcacattacccgcaacacaggtccatccagagagcgtctgaac	1511
Db	968	GCCCTTAACCTAAAGTTTCGGCACTTACCGCCACACGAGCCATCTCTGAAGGGGTTTCGGCG	1027
QY	1512	ggagagagcgcgtgcacatcatcagacctcgatcacatgatgaaagggcgtgcagtgccgaggc	1571
Db	1028	GCTCCGCCCGCTGACAGGTGTGAGACTTTCGGCATTAAGAGAGGGATGATGAGTGGCCCGCAC	1087
QY	1572	tcttcacatcctctgtctcccgcccgaggcgcccgcccaaggttgaggtcttaacgagctcg	1631
Db	1088	TTCCTCAGGCCCTTCGCGCTTCGCCGGCGGCTTCCTCTGTTCCGCTCACCGGCGTGC	1147
QY	1632	g-----g'gcgtccatctgaaagcgtcctgaaagccaaagggaaagcgcctctccagatt	1679
Db	1148	GCCCCCGCAGCGGAGAGAACGACGACGCCCTGAGAGAGTGGGGTGGAAAGTCCGCCAGT	1207
QY	1680	tcgcgcagacgcctgcgcgccttcgtaattctctgcgcgtgcgcg	1724
Db	1208	TCGGGCACACCATTCGCGGTGAGTTCCAGATACCGGGCTCTGTC	1252

RESULT 14
CEY39B6_3
WPCOMMENT
Sequence split into 4 fragments LOCUS Cey39B6 Accession Z95399

Continuation (4 of 4) of CEY39B6 from base 300001 (Z95399 *Caenorhabditis elegans* chromosome

Query Match	4.98;	Score 170.8;	DB 2;	Length 56870;
Best Local Similarity	48.18;	Prod No 3 30-13;		

[illegible][illegible]

RESULT 15
20017067

LOCUS	AB017067	83689 bp	DNA	PLNT	27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MUT20.				
ACCESSION	AB017067	BA0000015			
VERSION	AB017067.1	GI:3510343			
KEYWORDS					
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone:_lib:Mitsui P1				

ORGANISM	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (sites)
AUTHORS	Kanehara, T., Katoh, T., Sato, S., Nakamura, Y., Asamizu, E., Kotani, H.,
TITLE	Myajima, N. and Tabata, S.
	Structural analysis of Arabidopsis thaliana chromosome 5. IX.

JOURNAL DNA Res. 6 (3), 183-195 (1999)

REFERENCE 2 (bases 1 to 83669)
AUTHORS Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) Yasukazu Nakamura, Kazusa DNA Research

COMMENT

The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/Graal-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>),

FEATURES

SOURCE

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://gremlini.zool.iastate.edu/cgi-bin/sp.cgi>) Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is K16122 and the 3' clone is K5014.

Location/Qualifiers

```

1..83689
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MJC20"
/clone_id="Mitsui P1"
complement(join(393..464,554..643,717..822,901..981,
1054..1265))
/note="gene_id:MJC20.1
similar to unknown protein
sp145966"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08424.1"
/db_xref="GI:9757936"
/translation="MDOTONTAGIGINGSTIKETINATATVDSNEKINOTIINLD
GKNDERTKVLNKNCTAKNQVTKGSKDARKLHLELEAFPEDEVDRIDIRAS
YAAAKRLAQTOVLPNGDAKVKSEL"
2215..3432
/note="gene_id:MJC20.2"
/codon_start=1
/evidence=not_experimental
/product="SCARECROW gene regulator-like protein"
/protein_id="BAB08425.1"
/db_xref="GI:9757937"
/translation="MTYKRIDRLPSSDDPSAKRRIEPPEETLENDGAAILLSLL
LOCAEYATDHLREASTLSEISCEPSSPEPVAVAPALOTRIVISSYSGAS
PLSEKPLTVVOSOKIFSALOTYNSVPLIKFSHTNOAIFQALDGEDSHIIDLVN
OGIOWPALPHILASRPKLRISIRITCGSSDLSLSTGRAPASINIPPEPHIE
GITGNLIDPSQLATROGEAVVHMOHRLDVGNNLETLLEIRLKPMLITVBEEL
SYDGGSGFLGRFEALHYSALEFDALGDGEGSGERFVEQLTGLTKNIVAHGG
RRKMRKKEELSRGFRPVSIRGNPATQALLGLMGMVNGYTLVEENGTLRLGMKDL
LTPASAKSQPPD"
4272..4568
/note="gene_id:MJC20.3
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08426.1"
/db_xref="GI:9757938"
/translation="MMFRLRLTKVFRSPILACKWIVFADGSSGGVGRNGYIGLWS
SAVPSWGLAVATVMAVAFTVYSRRTYHNDLIVLIDFKAKKOCLETT"
join(5257..5345,5552..5687,5778..5827,6222..6275,
6493..7684)
/note="gene_id:MJC20.4"
/codon_start=1
/evidence=not_experimental
/product="GTPase activator protein of Rab-like small
GTPases-like protein"
/protein_id="BAB08427.1"
/db_xref="GI:9757939"
/translation="MIFCLLVLLIGISWTVCSDDEVYKSLRGSIGSPWLSRRKR
VLQPKQNAFTTEBGRISDGVKFLKKVRSQVHPSIREFVWFLGVLDLKNKEER
DSIRQLULEYENLRQCRERHERYENCDSKETAQSDNTEDSVLESHDIEVESST
RTITVDESKELNESIMQDENCEKSDVETEDAGNDSTNPETETSPFLAEVEES
HNTVNOKEISSPSKPSQADEDEFTWQRIIRLDVANDENVYSPTOAVSET
KARGLATOVGLNDYDLPCRFIHAARLVGLIAYAVOPEIGYCOGMSDLSPLIAY
MEDVLAIFMCFVGFMSKARHNPRLDEVGIRQISMSKIIFKFDILYHLEMLBED
CFYVYRVVYVLFRELTFEOTLCLMEYMAADQAIRTGAKATWGRIRLRAPTEDL
LYAIAASVIAORRRTIIEKYSMDIEMKECNSMAGHLDVWKLLDDADHLYVNHDKI"

```

CDS

```

join(8058..8569,8692..8811,9050..9119,9477..9590,
9713..9796,9924..9974,10134..10163,10275..10337,
10430..10495,10589..10717,10851..10886,10966..11085,
11163..11354,11483..11593)
/note="gb|AA032311.1
gene_id:MJC20.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08428.1"
/db_xref="GI:9757940"
/translation="MADVEEPQLONGAAPAESETEONPIPEPOLQTEBPKLTGELPEI
EADLIPPEVQSEVTDKAPKEVQSEVKEPEVTVYTDAPPEAQSEVKEPEVQSVTPT
KPDLLVDLSPGSEIEPIRSTVEQSESTVSLKDDGDKFTMRBELLSLSEEGD
GPHSSAPSESESQAPENNPMADLNRLOVNDDEGRSHORVLAFAKRYASAIEER
NPDHDLALYNMALILOESADNVPSPDSVSDLLLEACRYDEATRLCPLDYAYN
WALAISDRAKITGRTRKEADELMEQADNTERAKVOLNMSQALNMKGVLQDELQIYD
AREKERVTRATISKFRALRIQDFDHRATYMDVGLVLAEDTLTGSGSGKDMPEG
ELYSQSAIYIAAASHLAKPSYSVSSALRLVSMPLPLKGVLTAPVGNLSLPHSD
WKTEFEELNHERLOVLKPEPEREMGRNLSGKAEIEMTVERTKVTNITEIVSTPCA
DLTLPAGALCIDTIHGVFLVADSMESLDGMLAIRLVYTIYARGSVDVLAIIYG"
join(11993..12130,12206..12427,12522..12601,12703..12778,
12871..12961,13075..13101)
/note="gene_id:MJC20.6
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08429.1"
/db_xref="GI:9757941"
/translation="WASLSCLPICGLSSASLYTKARLLSSSSSPSPFLAYPLQNP
KILCKIERVITRAEFSDAPLVTAIGACILISFVFPAPKRVNDEEENSAIVTD
MLAAMGIISFIPFENMLSWFAMLDTRKSRVAYVALYVLYLSNLSIPESMWLP
ITSYIGIITHVOLDEASIANGVETLAFPRNTSDDESSKRIHKKHPRKGNSDEN"
join(13512..13697,13866..14012,14098..14196,14285..14476,
14612..14731,14818..14869,14970..15194)
/note="gene_id:MJC20.7"
/codon_start=1
/evidence=not_experimental
/product="GAMM1 protein-like"
/protein_id="BAB08430.1"
/db_xref="GI:9757942"
/translation="MASSSVRVYSTATSPSPSTSYKVGCTHNGSFHDEALCFMI
RLVDFSGADITRSRDKPRLAEBLDVLVGVGYDPEHDIRHHQGFEEVGHGPNTR
LSSAGLVYKHFEKELIAKLELVNDGDPVLRPLFVAYKSFMEALDAVDANGINRDYDQ
PFRVYNNHLSERVRGLNDWIDPQOSCEKNEFAORMALAGFELESVOFHARSDL
PARSIYMOCLEERFKTDPDSEIMILDRCPMKLHFLEQCKIEPLIKYIYODERA
KOMRYOAVAVAPDRENNRPLPEKMRGLRDELSKAATIPCCVFVHMSGFICGNQSD
GALSMAGTALRT"
15998..17122
/note="emb|CA072466.1
gene_id:MJC20.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08431.1"
/db_xref="GI:9757943"
/translation="MEISGEEDKEEAVTLPREVSKISISDGNKEFVYQILANGPNEOCE
NFRMDKPVYKLCDLQTRGLRHTNRKILIAOLAIFLITGHNLRFAVOLECYSG
EITSRFPNNVNAIVATISKDFEOPNSNTDLENDPDKDGVGVDSDHPIPVMGVDE
OGFPRNGCLITONVLAASSFDLRPNRYLACBEGSASQOVLNMLHTRPRKLYOPQK
YIIVDNKYNLPGFLAPYHGVSTNSREKAKFMPDRHLLHRLAIRHTFGALKEKFPIL
LSAPPYLOTVKLVIAACALHNVLRKPPDLVFRMEDEETLAEAGEDEBRVALLEEBO
VIVGOEHGPREPEVEDSLRLDEIASLMMHYVQNMST"
complement(join(17318..17892,17980..18361,18437..18594,
18670..18890,18991..19218,19303..19340,19706..19770,
19981..20140))
/note="gene_id:MJC20.9"
/codon_start=1
/evidence=not_experimental
/product="MAP kinase"
/protein_id="BAB08432.1"
/db_xref="GI:9757944"
/translation="MEGFNRRIKERRNHDRGRDRREIGCKSEMGIISFAKLRFNRSPRT

```

CDS

```

join(8058..8569,8692..8811,9050..9119,9477..9590,
9713..9796,9924..9974,10134..10163,10275..10337,
10430..10495,10589..10717,10851..10886,10966..11085,
11163..11354,11483..11593)
/note="gb|AA032311.1
gene_id:MJC20.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08428.1"
/db_xref="GI:9757940"
/translation="MADVEEPQLONGAAPAESETEONPIPEPOLQTEBPKLTGELPEI
EADLIPPEVQSEVTDKAPKEVQSEVKEPEVTVYTDAPPEAQSEVKEPEVQSVTPT
KPDLLVDLSPGSEIEPIRSTVEQSESTVSLKDDGDKFTMRBELLSLSEEGD
GPHSSAPSESESQAPENNPMADLNRLOVNDDEGRSHORVLAFAKRYASAIEER
NPDHDLALYNMALILOESADNVPSPDSVSDLLLEACRYDEATRLCPLDYAYN
WALAISDRAKITGRTRKEADELMEQADNTERAKVOLNMSQALNMKGVLQDELQIYD
AREKERVTRATISKFRALRIQDFDHRATYMDVGLVLAEDTLTGSGSGKDMPEG
ELYSQSAIYIAAASHLAKPSYSVSSALRLVSMPLPLKGVLTAPVGNLSLPHSD
WKTEFEELNHERLOVLKPEPEREMGRNLSGKAEIEMTVERTKVTNITEIVSTPCA
DLTLPAGALCIDTIHGVFLVADSMESLDGMLAIRLVYTIYARGSVDVLAIIYG"
join(11993..12130,12206..12427,12522..12601,12703..12778,
12871..12961,13075..13101)
/note="gene_id:MJC20.6
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08429.1"
/db_xref="GI:9757941"
/translation="WASLSCLPICGLSSASLYTKARLLSSSSSPSPFLAYPLQNP
KILCKIERVITRAEFSDAPLVTAIGACILISFVFPAPKRVNDEEENSAIVTD
MLAAMGIISFIPFENMLSWFAMLDTRKSRVAYVALYVLYLSNLSIPESMWLP
ITSYIGIITHVOLDEASIANGVETLAFPRNTSDDESSKRIHKKHPRKGNSDEN"
join(13512..13697,13866..14012,14098..14196,14285..14476,
14612..14731,14818..14869,14970..15194)
/note="gene_id:MJC20.7"
/codon_start=1
/evidence=not_experimental
/product="GAMM1 protein-like"
/protein_id="BAB08430.1"
/db_xref="GI:9757942"
/translation="MASSSVRVYSTATSPSPSTSYKVGCTHNGSFHDEALCFMI
RLVDFSGADITRSRDKPRLAEBLDVLVGVGYDPEHDIRHHQGFEEVGHGPNTR
LSSAGLVYKHFEKELIAKLELVNDGDPVLRPLFVAYKSFMEALDAVDANGINRDYDQ
PFRVYNNHLSERVRGLNDWIDPQOSCEKNEFAORMALAGFELESVOFHARSDL
PARSIYMOCLEERFKTDPDSEIMILDRCPMKLHFLEQCKIEPLIKYIYODERA
KOMRYOAVAVAPDRENNRPLPEKMRGLRDELSKAATIPCCVFVHMSGFICGNQSD
GALSMAGTALRT"
15998..17122
/note="emb|CA072466.1
gene_id:MJC20.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08431.1"
/db_xref="GI:9757943"
/translation="MEISGEEDKEEAVTLPREVSKISISDGNKEFVYQILANGPNEOCE
NFRMDKPVYKLCDLQTRGLRHTNRKILIAOLAIFLITGHNLRFAVOLECYSG
EITSRFPNNVNAIVATISKDFEOPNSNTDLENDPDKDGVGVDSDHPIPVMGVDE
OGFPRNGCLITONVLAASSFDLRPNRYLACBEGSASQOVLNMLHTRPRKLYOPQK
YIIVDNKYNLPGFLAPYHGVSTNSREKAKFMPDRHLLHRLAIRHTFGALKEKFPIL
LSAPPYLOTVKLVIAACALHNVLRKPPDLVFRMEDEETLAEAGEDEBRVALLEEBO
VIVGOEHGPREPEVEDSLRLDEIASLMMHYVQNMST"
complement(join(17318..17892,17980..18361,18437..18594,
18670..18890,18991..19218,19303..19340,19706..19770,
19981..20140))
/note="gene_id:MJC20.9"
/codon_start=1
/evidence=not_experimental
/product="MAP kinase"
/protein_id="BAB08432.1"
/db_xref="GI:9757944"
/translation="MEGFNRRIKERRNHDRGRDRREIGCKSEMGIISFAKLRFNRSPRT

```

CDS

```

join(8058..8569,8692..8811,9050..9119,9477..9590,
9713..9796,9924..9974,10134..10163,10275..10337,
10430..10495,10589..10717,10851..10886,10966..11085,
11163..11354,11483..11593)
/note="gb|AA032311.1
gene_id:MJC20.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08428.1"
/db_xref="GI:9757940"
/translation="MADVEEPQLONGAAPAESETEONPIPEPOLQTEBPKLTGELPEI
EADLIPPEVQSEVTDKAPKEVQSEVKEPEVTVYTDAPPEAQSEVKEPEVQSVTPT
KPDLLVDLSPGSEIEPIRSTVEQSESTVSLKDDGDKFTMRBELLSLSEEGD
GPHSSAPSESESQAPENNPMADLNRLOVNDDEGRSHORVLAFAKRYASAIEER
NPDHDLALYNMALILOESADNVPSPDSVSDLLLEACRYDEATRLCPLDYAYN
WALAISDRAKITGRTRKEADELMEQADNTERAKVOLNMSQALNMKGVLQDELQIYD
AREKERVTRATISKFRALRIQDFDHRATYMDVGLVLAEDTLTGSGSGKDMPEG
ELYSQSAIYIAAASHLAKPSYSVSSALRLVSMPLPLKGVLTAPVGNLSLPHSD
WKTEFEELNHERLOVLKPEPEREMGRNLSGKAEIEMTVERTKVTNITEIVSTPCA
DLTLPAGALCIDTIHGVFLVADSMESLDGMLAIRLVYTIYARGSVDVLAIIYG"
join(11993..12130,12206..12427,12522..12601,12703..12778,
12871..12961,13075..13101)
/note="gene_id:MJC20.6
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08429.1"
/db_xref="GI:9757941"
/translation="WASLSCLPICGLSSASLYTKARLLSSSSSPSPFLAYPLQNP
KILCKIERVITRAEFSDAPLVTAIGACILISFVFPAPKRVNDEEENSAIVTD
MLAAMGIISFIPFENMLSWFAMLDTRKSRVAYVALYVLYLSNLSIPESMWLP
ITSYIGIITHVOLDEASIANGVETLAFPRNTSDDESSKRIHKKHPRKGNSDEN"
join(13512..13697,13866..14012,14098..14196,14285..14476,
14612..14731,14818..14869,14970..15194)
/note="gene_id:MJC20.7"
/codon_start=1
/evidence=not_experimental
/product="GAMM1 protein-like"
/protein_id="BAB08430.1"
/db_xref="GI:9757942"
/translation="MASSSVRVYSTATSPSPSTSYKVGCTHNGSFHDEALCFMI
RLVDFSGADITRSRDKPRLAEBLDVLVGVGYDPEHDIRHHQGFEEVGHGPNTR
LSSAGLVYKHFEKELIAKLELVNDGDPVLRPLFVAYKSFMEALDAVDANGINRDYDQ
PFRVYNNHLSERVRGLNDWIDPQOSCEKNEFAORMALAGFELESVOFHARSDL
PARSIYMOCLEERFKTDPDSEIMILDRCPMKLHFLEQCKIEPLIKYIYODERA
KOMRYOAVAVAPDRENNRPLPEKMRGLRDELSKAATIPCCVFVHMSGFICGNQSD
GALSMAGTALRT"
15998..17122
/note="emb|CA072466.1
gene_id:MJC20.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08431.1"
/db_xref="GI:9757943"
/translation="MEISGEEDKEEAVTLPREVSKISISDGNKEFVYQILANGPNEOCE
NFRMDKPVYKLCDLQTRGLRHTNRKILIAOLAIFLITGHNLRFAVOLECYSG
EITSRFPNNVNAIVATISKDFEOPNSNTDLENDPDKDGVGVDSDHPIPVMGVDE
OGFPRNGCLITONVLAASSFDLRPNRYLACBEGSASQOVLNMLHTRPRKLYOPQK
YIIVDNKYNLPGFLAPYHGVSTNSREKAKFMPDRHLLHRLAIRHTFGALKEKFPIL
LSAPPYLOTVKLVIAACALHNVLRKPPDLVFRMEDEETLAEAGEDEBRVALLEEBO
VIVGOEHGPREPEVEDSLRLDEIASLMMHYVQNMST"
complement(join(17318..17892,17980..18361,18437..18594,
18670..18890,18991..19218,19303..19340,19706..19770,
19981..20140))
/note="gene_id:MJC20.9"
/codon_start=1
/evidence=not_experimental
/product="MAP kinase"
/protein_id="BAB08432.1"
/db_xref="GI:9757944"
/translation="MEGFNRRIKERRNHDRGRDRREIGCKSEMGIISFAKLRFNRSPRT

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2002, 15:36:30 ; Search time 298.88 Seconds

(without alignments)
10068.294 Million cell updates/sec

Title: US-09-265-585c-95

Perfect score: 3510

Sequence: 1 ccgcctagctcagcctactca.....tagcctcaagcaagtattta 3510

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues 1861242

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1980.DAT:*
2:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1981.DAT:*
3:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1982.DAT:*
4:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1983.DAT:*
5:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1984.DAT:*
6:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1985.DAT:*
7:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1986.DAT:*
8:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1987.DAT:*
9:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1988.DAT:*
10:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1989.DAT:*
11:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1990.DAT:*
12:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1991.DAT:*
13:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1992.DAT:*
14:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1993.DAT:*
15:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1994.DAT:*
16:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1995.DAT:*
17:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1996.DAT:*
18:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1997.DAT:*
19:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1998.DAT:*
20:	/SIDS2/gcgcdata/geneseq/geneseqn/NA1999.DAT:*
21:	/SIDS2/gcgcdata/geneseq/geneseqn/NA2000.DAT:*
22:	/SIDS2/gcgcdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3510	100.0	3510	21	Maize Scarecrow nu
2	698.4	19.9	2151	18	Maize ZCARECROW ZC
3	698.4	19.9	2151	21	Maize ZCR gene par
4	358.4	10.2	2163	18	Arabidopsis SCAREC
5	262.4	7.5	2161	21	Arabidopsis SCAREC
6	218.4	6.2	969	21	Maize CBPBT14 par
7	172.4	4.9	2709	20	Composite DNA sequ
8	171.8	4.9	1768	20	Wheat Rht clone C1
9	171.8	4.9	2125	20	Wheat Rht clone 5a
10	169.4	4.8	1369	21	Arabidopsis thaliana
11	169	4.8	1746	20	Consensus cDNA seq

12	157.4	4.5	1328	21	AAC39840
13	155.8	4.4	2255	20	AAK36280
14	140.2	4.0	1337	20	AAZ17263
15	131.2	3.7	114955	20	AAK53491
16	130	3.7	321	18	AAT95765
17	130	3.7	321	21	AAC65296
18	127.4	3.6	377	20	AAK36268
19	122.4	3.5	332	20	AAK36255
20	118	3.4	1000	21	AAA02484
21	116.2	3.3	114955	20	AAK53491
22	108.6	3.1	1218	21	AAA02488
23	107.8	3.1	3957	22	AAA09686
24	106.8	3.0	1593	21	AAA02504
25	106	3.0	8438	15	AAQ73500
26	101.8	2.9	1459	21	AAQ02528
27	99.8	2.8	1359	20	AAZ17254
28	98	2.8	1127	21	AAA02477
29	96.8	2.8	100	21	AAC65284
30	95.2	2.7	100	18	AAT95769
31	94.4	2.7	1779	22	AAF25480
32	94.4	2.7	1779	22	AAF25481
33	93.2	2.7	1523	20	AAZ17507
34	93.2	2.7	1523	20	AAZ17479
35	90	2.6	1925	20	AAK90924
36	90	2.6	5120	22	AAC84677
37	89.4	2.5	320	21	AAA38183
38	88.8	2.5	1126	21	AAA02538
39	88.4	2.5	1126	21	AAA02538
40	88.4	2.5	6530	14	AAO51557
41	88.4	2.5	6530	20	AAZ22072
42	88.4	2.5	6530	22	AAC68948
43	87.2	2.5	1926	21	AAA50254
44	87.2	2.5	1926	22	AAF82902
45	87.2	2.5	2580	21	AAAT5454

ALIGNMENTS

RESULT 1

AAAC65299	standard: cDNA, 3510 BP.
ID	AAAC65299
AC	AAAC65299;
XX	
DT	09-FEB-2001 (first entry)
XX	
DE	Maize Scarecrow nucleotide sequence.
XX	
KW	Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
KW	transgenic plant; cell division; molecular marker; herbicide resistance;
KW	salt resistance; pathogen resistance; insect resistance; ss.
XX	
OS	Zea mays.
XX	
PN	W0200053723-A2.
XX	
PD	14-SEP-2000.
XX	
PF	07-MAR-2000; 2000WO-US05875.
XX	
PR	10-MAR-1999; 99US-0265585.
XX	
PA	(UYNV) UNIV NEW YORK STATE.
XX	
PI	Bentley PN, Di Laurenzio L, Mysoska-Diller J, Malamy JB, Pysh L,
PI	Helariutta Y, Bruce W, Lim J;
XX	
DR	WPI; 2000-594315/56.
XX	
P	P-PSDB; AAB28595.
PT	Scarecrow gene useful for producing transgenic plants expressing genes
PT	whose product increases starch, lignin or cellulose biosynthesis and

PT confers herbicide, pathogen or insect resistance
XX
PS Disclosure, Fig 25; 200pp; English.

The present sequence is given in a specification relating to the structure and function of a regulatory gene, SCARCEROW (SCR). SCARCEROW-like (SCL) genes encoding proteins containing an amino acid sequence similar to the sequence of MOTIF III (VIII) of Arabidopsis SCR protein are disclosed. SCR and SCL genes are useful for producing transgenic plants whose cell division is modified and root and/or stem development and gravitropism of stem or hypocotyl is altered. Cell division is increased or decreased in roots resulting in thicker or thinner root development. The transgenic plants are useful for expressing a gene of interest encoding a gene product that confers herbicide, salt, pathogen or insect resistance in root or embryos and genes encoding starch, lignin or cellulose biosynthesis in shoots. The SCR gene also confers less susceptibility to lodging in the transgenic plants than a wild-type plant. SCR gene sequences are also useful as molecular markers for a quantitative trait e.g. root or gravitropism trait in molecular breeding of crop plants.

Sequence 3510 BP; 651 A; 1225 C; 946 G; 688 T; 0 other;

[illegible][illegible]

Dh	1741	tgaccggagagaagctaaagggttcaagagcgaggcgctgcgcgtccacgtgctgcaca	1800
Oy	1801	ctcgctctacgaagctcaactgctgcgtcgcacccaacagcgtctggtccatcccaaaagttaga	1860
Dh	1801	ctcgctctacgaagctcaactgctgcgtcgcacccaacagcgtctggtccatcccaaaagttaga	1860
Oy	1861	aggagttacaacatctctcgactctgactcttccttgctacacgtatccaactcttgatgacac	1920
Dh	1861	aggagttacaacatctctcgactctgactcttccttgctacacgtatccaactcttgatgacac	1920
Oy	1921	atggcgcacttttgagctgactacaacatttgatttagccaatgagacatccagttacaactaa	1980
Dh	1921	atggcgcacttttgagctgactacaacatttgatttagccaatgagacatccagttacaactaa	1980
Oy	1981	tcctaaaaataaactcttcagagagtcttcctcagttaaaaaacgcgcttttttggagctcaaa	2040
Dh	1981	tcctaaaaataaactcttcagagagtcttcctcagttaaaaaacgcgcttttttggagctcaaa	2040
Oy	2041	agcttgcattatgacccaacaaacttctcaggttataaaaaaggttgaactctgacatgct	2100
Dh	2041	agcttgcattatgacccaacaaacttctcaggttataaaaaaggttgaactctgacatgct	2100
Oy	2101	tttgagacgcatgctgctactctgaagtactggtatagaagagatctccatgctcctttgtta	2160
Dh	2101	tttgagacgcatgctgctactctgaagtactggtatagaagagatctccatgctcctttgtta	2160
Oy	2161	tcccagagaccacaagaatagtaatttcgcgtccatttggtagaggttctgtatca	2220
Dh	2161	tcccagagaccacaagaatagtaatttcgcgtccatttggtagaggttctgtgtatca	2220
Oy	2221	tcacttggagagatgagctgtaaaagtctcctatcatcatgacccaactcttccatgycanaag	2280
Dh	2221	tcacttggagagatgagctgtaaaagtctcctatcatcatgacccaactcttccatgycanaag	2280
Oy	2281	tttttaatttccaagttggcagagagatgattactggtgactgagatgagagtgtaactgtaac	2340
Dh	2281	tttttaatttccaagttggcagagagatgattactggtgactgagatgagagtgtaactgtaac	2340
Oy	2341	ttctacacctaacaacatcaattatataataaatacatcatccggagagctaaactccagaa	2400
Dh	2341	ttctacacctaacaacatcaattatataataaatacatcatccggagagctaaactccagaa	2400
Oy	2401	aggcttaatcaaaagttgcaatcttctcnaaaygctgycatatgccaatgycnaatggtacatg	2460
Dh	2401	aggcttaatcaaaagttgcaatcttctcnaaaygctgycatatgccaatgycnaatggtacatg	2460
Oy	2461	ggcactgtacttttttcggtggaacacatctcgtctccaagatgagatgaaagccgtcaactgc	2520
Dh	2461	ggcactgtacttttttcggtggaacacatctcgtctccaagatgagatgaaagccgtcaactgc	2520
Oy	2521	aatctccttatattgtagacaacattatttagaagaagaacgcttaagagatgagccggtccgcgc	2580
Dh	2521	aatctccttatattgtagacaacattatttagaagaagaacgcttaagagatgagccggtccgcgc	2580
Oy	2581	aagggcaatgcatatcgttcaagacagagcgaggttcggaatcgatggtctagctcggtgac	2640
Dh	2581	aagggcaatgcatatcgttcaagacagagcgaggttcggaatcgatggtctagctcggtgac	2640
Oy	2641	aaatcgcagcagggatataataaatacatattgtgcatattgcatccatccccaacaactgca	2700
Dh	2641	aaatcgcagcagggatataataaatacatattgtgcatattgcatccatccccaacaactgca	2700
Oy	2701	ggctcgcccccaaggttgggtgacaatgttggaacagagactctgacacatcgcggtccctcc	2760
Dh	2701	ggctcgcccccaaggttgggtgacaatgttggaacagagactctgacacatcgcggtccctcc	2760
Oy	2761	ttggcgcgcttcgttggagggccatcccaactaactactcgcgtcttgcactcgcgttgacgcga	2820
Dh	2761	ttggcgcgcttcgttggagggccatcccaactaactactcgcgtcttgcactcgcgttgacgcga	2820
Oy	2821	gtctacggtcgagagacgcccccgagcgacgctgtgtggaacagacgtgtgtcgcgggaga	2880
Dh	2821	gtctacggtcgagagacgcccccgagcgacgctgtgtggaacagacgtgtgtcgcgggaga	2880

[illegible]

CC confers less susceptibility to lodging in the transgenic plants than a
CC wild-type plant. SCR gene sequences are also useful as molecular markers
CC for a quantitative trait e.g. root or gravitropism trait in molecular
CC breeding of crop plants.

SQ Sequence 2161 BP; 569 A; 668 C; 312 G; 606 T; 6 other;

Query Match	7.58;	Score 262.4;	DB 21;	Length 2161;
Best Local Similarity	55.08;	Pred. No. 4.9e-34;		
Matches 610;	Conservative 0;	Mismatches 477;	Indels 22;	Gaps 4

[illegible]

Db	1394	gctctcccaattacacagataaaccttccctcccttcttaacctctgccttaccgtgag	1453
QY	1727	aagcgcgcgaatgcttgaccgcggagaagactgagggttaccagcgaggagccgtgcgcgc	1786
Db	1454	aaagtctgaaccttcgcactgcactgagactaactcgtacgaataaccacaactctccctt	1513
QY	1787	cactgcctgcacccactctgcctcagcagctgaactgcctgcctgcactccaagctctgcgc	1846
Db	1514	cactgccttcaacactctctcttaatgatcactccctctgatgcacaaactctctcgtta	1573
QY	1847	atccaaagtgtagagagtgaaacacatc	1875
Db	1574	ctccaaacgctaaataaacattccctt	1602

RESULT 6

ID AAC65327 standard; cDNA; 969 BP.

AC AAC65327

DT 09-FEB-2001 (first entry)

DE Maize CBPBT44 partial cDNA sequence:

SCARECROW gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;
transgenic plant; cell division; molecular marker; herbicide resistance;
salt resistance; pathogen resistance; insect resistance; ss.

OS Zea mays.

PN W0200053723-A2

PD 14-SEP-2000.

PF 07-MAR-2000; 2000WO-US05875.

PR 10-MAR-1999; 99US-0265585.

PA (UYN) UNIV NEW YORK STATE.

PI Benfey PN, Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L;

XX

DR P-PSDB; AAB28603

PT Scarecrow gene useful for producing transgenic plants expressing genes

PT confers herbicide, pathogen or insect resistance

PS Example 8; Fig 33; 200pp; English.

The present sequence is given in a specification relating to the structure and function of a regulatory gene, SCARECROW (SCR). SCARECROW-like (SCL) genes encoding proteins containing an amino acid sequence similar to the sequence of MOTIF III (WHIT) of Arabidopsis SCR protein are disclosed. SCR and SCL genes are useful for producing transgenic plants whose cell division is modified and root and/or stem development and gravitropism of stem or hypocotyl is altered. Cell division is increased or decreased in roots resulting in thicker or thinner root development. The transgenic plants are useful for expressing a gene of interest encoding a gene product that confers herbicide, salt, pathogen or insect resistance in root or embryos and genes encoding starch, lignin or cellulose biosynthesis in shoots. The SCR gene also confers less susceptibility to lodging in the transgenic plants than a wild-type plant. SCR gene sequences are also useful as molecular markers for a quantitative trait e.g. root or gravitropism trait in molecular breeding of crop plants.

SQ Sequence 969 BP; 133 A; 333 C; 360 G; 143 T; 0 other;

Disclosure: Fig 3a; 88pp; English.

AA
AC
AAX36278;

XX	
DT	16-JUL-1999 (first entry)
XX	
DE	wheat Rht clone C15-1 cDNA sequence.
XX	
KM	Rht gene; homologue; Triticum aestivum; wheat; growth inhibition;
KW	antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
KW	pachlobutrazol, ss.
XX	
OS	Triticum aestivum.
XX	
FN	MO9909174-A1.
PD	
XX	25-FEB-1999.
XX	
EF	07-AUG-1998; 98WO-G802383.
XX	
PR	13-AUG-1997; 97GB-0017192.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PL	Harberd NF, Peng J, Richards DE;
XX	
DR	WPI; 1999-181040/15.
XX	
DR	P-PSDB; AAY02539.
XX	
PT	New Triticum Aestivum polynucleotides - encode a polypeptide which
PT	provides inhibition of the growth of plants, which inhibition is
XX	antagonised by gibberellin, used to confer a dwarf phenotype
XX	
PS	Disclosure; Fig 7a; 88pp; English.
XX	
CC	The specification describes polypeptides encoded by the Rht gene (and
CC	its homologues) that, when expressed in Triticum Aestivum, inhibit
CC	growth of the plant. This growth inhibition is antagonised by
CC	gibberellin. The products can be used to provide Rht expression in
CC	plants, conferring a dwarf phenotype on a plant which is correctable
CC	by treatment with gibberellin. In addition, the products can be
CC	used to produce Rht mutant plants which are dwarfed compared with
CC	wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC	may be made by knocking out Rht or the relevant homologous gene in
CC	the plant of interest. Plants may be made which are resistant to
CC	compounds which inhibit gibberellin biosynthesis, such as paclobutrazol
CC	e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC	dwarf but let crop plants grow tall. The present sequence represents
CC	the wheat Rht clone C15-1 cDNA sequence.
XX	
SQ	Sequence 1768 BP; 368 A; 595 C; 527 G; 278 T; 0 other;

[illegible][illegible]

RESULT 9

AAX36279
ID AAX36279 standard; DNA; 2125 BP.

AC AAX36279;

DT 16-JUL-1999 (first entry)

DE wheat Rht clone 5a1 genomic sequence

KW Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;

KW paclobutrazol; ss.

05 Triticum aestivum.

PN W09909174-A1.

PD 25-FEB-1999.

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

XX
XX

PT provides inhibition of the growth of plants, which inhibits

[illegible][illegible]

its homologues) that, when expressed in *Triticum aestivum*, inhibit

gibberellin. The products can be used to provide Rht expression in

by treatment with gibberellin. In addition, the products can be

CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants can be made by knocking out part of the relevant homologous gene in

CC the plant of interest. Plants may be made which are resistant to

01-JUL-1999;	PR	99US-0142154.
02-JUL-1999;	PR	99US-0142055.
06-JUL-1999;	PR	99US-0142390.
08-JUL-1999;	PR	99US-0142803.
09-JUL-1999;	PR	99US-0142920.
12-JUL-1999;	PR	99US-0142977.
13-JUL-1999;	PR	99US-0143542.
14-JUL-1999;	PR	99US-0143624.
15-JUL-1999;	PR	99US-0144005.
16-JUL-1999;	PR	99US-0144085.
16-JUL-1999;	PR	99US-0144086.
19-JUL-1999;	PR	99US-0144335.
19-JUL-1999;	PR	99US-0144334.
20-JUL-1999;	PR	99US-0144352.
20-JUL-1999;	PR	99US-0144632.
20-JUL-1999;	PR	99US-0144684.
21-JUL-1999;	PR	99US-0144814.
21-JUL-1999;	PR	99US-0145086.
21-JUL-1999;	PR	99US-0145088.
22-JUL-1999;	PR	99US-0145085.
22-JUL-1999;	PR	99US-0145087.
22-JUL-1999;	PR	99US-0145089.
22-JUL-1999;	PR	99US-0145192.
23-JUL-1999;	PR	99US-0145145.
23-JUL-1999;	PR	99US-0145218.
23-JUL-1999;	PR	99US-0145224.
26-JUL-1999;	PR	99US-0145276.
27-JUL-1999;	PR	99US-0145313.
27-JUL-1999;	PR	99US-0145318.
27-JUL-1999;	PR	99US-0145319.
28-JUL-1999;	PR	99US-0145351.
02-AUG-1999;	PR	99US-0146386.
02-AUG-1999;	PR	99US-0146388.
02-AUG-1999;	PR	99US-0146389.
03-AUG-1999;	PR	99US-0147308.
04-AUG-1999;	PR	99US-0147204.
04-AUG-1999;	PR	99US-0147302.
05-AUG-1999;	PR	99US-0147192.
05-AUG-1999;	PR	99US-0147260.
06-AUG-1999;	PR	99US-0147303.
06-AUG-1999;	PR	99US-0147416.
09-AUG-1999;	PR	99US-0147493.
09-AUG-1999;	PR	99US-0147935.
10-AUG-1999;	PR	99US-0148171.
11-AUG-1999;	PR	99US-0148319.
12-AUG-1999;	PR	99US-0148341.
13-AUG-1999;	PR	99US-0148565.
13-AUG-1999;	PR	99US-0148684.
16-AUG-1999;	PR	99US-0149368.
17-AUG-1999;	PR	99US-0149175.
18-AUG-1999;	PR	99US-0149426.
20-AUG-1999;	PR	99US-0149722.
20-AUG-1999;	PR	99US-0149723.
20-AUG-1999;	PR	99US-0149929.
23-AUG-1999;	PR	99US-0149902.
23-AUG-1999;	PR	99US-0149930.
25-AUG-1999;	PR	99US-0150566.
26-AUG-1999;	PR	99US-0150884.
27-AUG-1999;	PR	99US-0151065.
27-AUG-1999;	PR	99US-0151066.
27-AUG-1999;	PR	99US-0151080.
30-AUG-1999;	PR	99US-0151303.
31-AUG-1999;	PR	99US-0151438.
01-SEP-1999;	PR	99US-0151930.
07-SEP-1999;	PR	99US-0152363.
10-SEP-1999;	PR	99US-0153070.
13-SEP-1999;	PR	99US-0153758.
15-SEP-1999;	PR	99US-0154019.
16-SEP-1999;	PR	99US-0154039.
20-SEP-1999;	PR	99US-0154779.
22-SEP-1999;	PR	99US-0155139.
23-SEP-1999;	PR	99US-0155486.
24-SEP-1999;	PR	99US-0155659.
28-SEP-1999;	PR	99US-0156458.
29-SEP-1999;	PR	99US-0156596.
04-OCT-1999;	PR	99US-0157117.
05-OCT-1999;	PR	99US-0157753.
06-OCT-1999;	PR	99US-0157865.
07-OCT-1999;	PR	99US-0158029.
08-OCT-1999;	PR	99US-0158232.
12-OCT-1999;	PR	99US-0158369.
13-OCT-1999;	PR	99US-0159293.
13-OCT-1999;	PR	99US-0159294.
14-OCT-1999;	PR	99US-0159329.
14-OCT-1999;	PR	99US-0159330.
14-OCT-1999;	PR	99US-0159331.
14-OCT-1999;	PR	99US-0159637.
14-OCT-1999;	PR	99US-0159638.
18-OCT-1999;	PR	99US-0159584.
21-OCT-1999;	PR	99US-0160741.
21-OCT-1999;	PR	99US-0160767.
21-OCT-1999;	PR	99US-0160768.
21-OCT-1999;	PR	99US-0160770.
21-OCT-1999;	PR	99US-0160814.
21-OCT-1999;	PR	99US-0160815.
22-OCT-1999;	PR	99US-0160980.
22-OCT-1999;	PR	99US-0160981.
22-OCT-1999;	PR	99US-0160989.
25-OCT-1999;	PR	99US-0161404.
25-OCT-1999;	PR	99US-0161405.
25-OCT-1999;	PR	99US-0161406.
26-OCT-1999;	PR	99US-0161359.
26-OCT-1999;</		

QY	1602	gcccgcgccaggggtgaagatcaacggcctcgggggagctcagatggaagcgtcggagggccacgg	1661	
Db	533	aactccgatatcaatcgaatcaaccggatcttggtctcccccgcgatactacgtccgttcaactg	592	
QY	1662	ggaagcgctctccgatcttcgcgcgaacacgcctcggcctgaccttcgagttctcgccgcctg	1721	
Db	593	gcgcgagacgtcgcgagatcttcgcatactgtaaccctcccttcgattcaatccaatctg	652	
QY	1722	ccgaggaagcgccggaatgttgacccggaaaagctgaggtctacggagc---	ggagagccgg	1778
Db	653	aagagcaaatcggaaaacccaatctgatacggagccaaactcgcaacagaaagaggaagctg	712	
QY	1779	tcgcgcgtccacttgcgcgcacacactcgctctacgacgtcacttgcgcgcgactccaacacgg	1838	
Db	713	tgtgtgtcatctgatacgacgacccggtttatatgtatcgatccagggaacaatctcggagact	772	
QY	1839	tctgtgcctcatccaagaagtggaa	1861	
Db	773	tgaagattctacgagagctcgtaa	795	

RESULT 11

AAx36285
ID AAx36285 standard; cDNA; 1746 BP.

AC AAX36285;

DT 16-JUL-1999 (first entry)

Consensus cDNA sequence of wheat Rht clone C15-1.

Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;

paclobutrazol; ss.

Trillicum aestivum.

PN WO9909174-A1.

PD 25-FEB-1999

PF 07-AUG-1998; 98WO-GB02383.

PR 13-AUG-1997; 97GB-0017192.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Harberd NP, Peng J, Richards DE;

DR WPI; 1999-181040/15.

PT New Triticum Aestivum polynucleotides - encode a polypeptide which

PT antagonised by gibberellin, used to confer a dwarf phenotype

PS Disclosure; Fig 2a; 88pp; English.

CC The specification describes polypeptides encoded by the Rht gene (and

growth of the plant. This growth inhibition is antagonised by

plants, conferring a dwarf phenotype on a plant which is correctable

used to produce Rht mutant plants which are dwarfed compared with

may be made by knocking out Rht or the relevant homologous gene in

compounds which inhibit gibberellin biosynthesis, such as paclobutrazol

cc dwarf but let crop plants grow tall] The present sequence
cc e.g.: to allow use of gibberellin biosynthesis inhibitor to keep

XX represents the consensus cDNA sequence of wheat Rnt clone C15-1.

sequence 1/40 BP; 3/3 A; 5/0 C; 4/5 G; 2/8 T; 8 other;

Query Match	4.8%;	Score 169;	DB 20;	Length 1746;
Best Local Similarity	61.0%;	Pred. No. 5.8e-19;		
Matches 346;	Conservative	0;	Mismatches 185;	Indels 36;
				Gaps 3;

[illegible]

RESULT 12

AAC39840
ID AAC39840 standard; DNA; 1328 BP

AC AAC39840;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 26094.

KW Hybridisation assay; genetic mapping; gene expression control;

metabolic pathway; promoter; termination sequence; ss

OS *Arabidopsis thaliana*

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 09-MAR-1999: 99US-0123548.

PK 43-MAR-1999; 9905-0123/88.

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128274.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140941.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155489.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

PR	08-OCT-1999:	99US-0158232.
PR	12-OCT-1999:	99US-0158369.
PR	13-OCT-1999:	99US-0159293.
PR	13-OCT-1999:	99US-0159294.
PR	13-OCT-1999:	99US-0159295.
PR	14-OCT-1999:	99US-0159329.
PR	14-OCT-1999:	99US-0159330.
PR	14-OCT-1999:	99US-0159331.
PR	14-OCT-1999:	99US-0159637.
PR	14-OCT-1999:	99US-0159638.
PR	18-OCT-1999:	99US-0159584.
PR	21-OCT-1999:	99US-0160741.
PR	21-OCT-1999:	99US-0160767.
PR	21-OCT-1999:	99US-0160768.
PR	21-OCT-1999:	99US-0160770.
PR	21-OCT-1999:	99US-0160814.
PR	21-OCT-1999:	99US-0160815.
PR	22-OCT-1999:	99US-0160980.
PR	22-OCT-1999:	99US-0160981.
PR	22-OCT-1999:	99US-0160989.
PR	25-OCT-1999:	99US-0161404.
PR	25-OCT-1999:	99US-0161405.
PR	25-OCT-1999:	99US-0161406.
PR	26-OCT-1999:	99US-0161359.
PR	26-OCT-1999:	99US-0161360.
PR	26-OCT-1999:	99US-0161361.
PR	28-OCT-1999:	99US-0161920.
PR	28-OCT-1999:	99US-0161992.
PR	28-OCT-1999:	99US-0161993.
PR	29-OCT-1999:	99US-0162142.

Query Match	4.58;	Score 157.4;	DB 21;	Length 1328;
Best Local Similarity	53.98;	Pred. No. 4.3e-17;		
Matches 368;	Conservative 0;	Mismatches 31;	Indels 4;	Gaps 2;

[illegible]

Accession	Sequence	Position
D6	aagcgcataatcgcgaaccttaatcgtatgcgagtcgaactctgcgaacgcgacgaagcggaagctgc	670
QY	tcgcgcgtccactatgctctgcacacactgcgtctctagaagctcaactgtctctgcattccacaacgcg	1833
D6	tcgaggtctcatctgcgaacgcacccggttatatgattgttaacgaggaacgactctcgsagact	730
QY	tcgtgctccatccaaaggtagaa	1861
D6	tcgagattctacgagagctgaa	731

RESULT	13
AA36280	
ID	AA36280 standard; DNA; 2255 BP

16-JUL-1999 (first entry)
Maize 1a1 genomic clone sequence.

KW Rht gene; homologue; *Triticum aestivum*; wheat; growth inhibition;
 KW antagonist; gibberellin; dwarf phenotype; gibberellin biosynthesis;
 KW paclobutrazol; maize; ss.

Zea mays.

W09909174-A1.

25-FEB-1999.

07-AUG-1998; 98WO-GB02383.

13-AUG-1997; 97GB-0017192

(PLAN-) PLANT BIOSCIENCE LTD.

Harberd NP, Peng J, Richards DE;

WPT; 1999-181040/15.

P-PSDB; AAY02541.

PT New *Triticum aestivum* polynucleotides - encode a polypeptide which provides inhibition of the growth of plants, which inhibition is antagonised by gibberellin, used to confer a dwarf phenotype

Disclosure; Fig 9a; 88pp; English.

AA The specification describes polypeptides encoded by the Rht gene (and
CC its homologues) that, when expressed in *Triticum aestivum*, inhibit
CC growth of the plant. This growth inhibition is antagonised by
CC gibberellin. The products can be used to provide Rht expression in
CC plants, conferring a dwarf phenotype on a plant which is correctable
CC by treatment with gibberellin. In addition, the products can be
CC used to produce Rht mutant plants which are dwarfed compared with
CC wild-type, the dwarfing being gibberellin-insensitive. Taller plants
CC may be made by knocking out Rht or the relevant homologous gene in
CC the plant of interest. Plants may be made which are resistant to
CC compounds which inhibit gibberellin biosynthesis, such as paclobutrazol
CC e.g. to allow use of a gibberellin biosynthesis inhibitor to keep weeds
CC dwarf but let crop plants grow tall. The present sequence represents
CC the maize 1a1 genomic clone sequence.

Sequence 2255 BP; 334 A; 817 C; 737 G; 367 T; 0 other;

Query Match	4.4%;	Score 155.8;	DB 20;	Length 2255;
Best Local Similarity	56.7%;	Pred. No. 8e-17;		
Matches 370; Conservative	0;	Mismatches 247;	Indels 36;	Gaps 3;

QY 1086 ccgcagcgcgcgcgcgcacaaagcagctctgcgcgcgcgaagagcgaag 1145
||| | | | | | | | | | | | | | | | |
Db 822 cccttctcgaagctctgcgcgcgcgcgcgaagcattccgcgcgcgcgaacgcgcgcgcg 881

Db 108278 VAGGCGAGCCAGGCGCCGGCCCCCGCCCNHNNHNSCCVAGGCGAGCCAGGCCCCCGCC 108219
QY 838 gctccttggagctccgctccctcctcctcctcagcagaccgcccacatgcccgcgc 897
Db 108218 CCCCCCNHNNHNSCCVAGGCGAGGCGCGCGCCGCCGCCCNHNNHNSGCCVAG 108159
QY 898 gccgcagccgcagcagcatgctctcttgcagggcgtccgcgcgcgtccgcggggct 957
Db 108158 GCGAGCCAGGCGCGCGCGCCGCCCNHNNHNSGCCVAGGCGAGCGCGC-CGCCG 108100
QY 958 gacgtccctcccccgcacagcttcctgcagcaagccgcacagagcatccaccgctg 1017
Db 108099 CCCCCGCNNHNNHNSGGGCCCVAGGCGAGCCAGGCGCGCGCCGCCCNHNNHNSGG 108040
QY 1018 ccagcagcaacagcagagagacatccgcgcgcagtcgcaccagccccgcagcgc 1077
Db 108039 GCCCVAGGCGAGGCGCGCGCGCCGCCCNHNNHNSCCGCGCGGCGCNHNNHNSCC 107980
QY 1078 ggaagagaccgc 1137
Db 107979 GGGCCGGGCGNNHNNHNSCCGGCGCGGCGNNHNNHNSCCGGCGCGGCGNNHNNHNSCC 107920
QY 1138 gccgaaggaagagcagc 1197
Db 107919 GGGCCGGGCGGCGNNHNNHNSCCGGCGCGGCGCGCAHNNHNNHNSCCGGCGCGGCGCAA 107860
QY 1198 gctgctgcagtgccgc 1251
Db 107859 NNHNNHNSCCGGGCGCGGCGGCAAGNNHNNHNSCCGGGCGCGGCGCAAGCNHNNHNSCC 107800
QY 1252 gacgcgtctggaatcgc 1311
Db 107799 GGGCCGGGCGGCAAGCCNNHNNHNSCCGGCGCGGCGCAAGCCNNHNNHNSCCGGCGC 107740
QY 1312 cgcctacttcgc 1371
Db 107739 GGGCGGGGCAAGCGGNNHNNHNSCCGGGCGCGGCGGCAAGCGGNNHNNHNSCCGGGCGG 107680
QY 1372 gccgcgtgc 1431
Db 107679 GGGCGGGCAAGCGGNNHNNHNSCCGGGCGCGGCGGCAAGCGGCGGCGGCGGCGGCGG 107620
QY 1432 ccaggtgttcaagc 1491
Db 107619 CCGGGCGGCAAGCGG--GCCGNNHNNHNSCCGGGCGGCGGCAAGCGGCGGCGGCGGCGG 107563
QY 1492 catccaggaagcgttcgc 1551
Db 107562 HNNHNSCCGGGCGGCGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 107503
QY 1552 ggggcgtgcagtgccgc 1611
Db 107502 GGGCGCGGCGNNHNNHNSCCGGGCGGCGGCGGCAAGCGGCGGCGGCGGCGGCGGCGG 107443
QY 1612 ggtgaagctcaacgc 1671
Db 107442 GGGCGGCGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 107383
QY 1672 ctccgatttcgc 1731
Db 107382 GGCNNHNNHNSCCGGGCGGCGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 107323
QY 1732 cggcaatgttgaccgc 1791
Db 107322 CCGGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 107263
QY 1792 gctgcaccacgcgc 1805
Db 107262 GCGCGAHHNNHNSCC 107249

Search completed: April 22, 2002, 18:00:24
Job time: 8634 sec

Db 3928329 GTACACGGCTGCGGCGGCTGGCTAGCCGGTGTCTACAGCCGCCCTTTGCGCCCTTG6GCC 3928270
QY 1403 cgcctcacgagcgcgagtcgcgcgcgcgttcacaggtgttcaacgcatcagccctcgtc 1462
Db 3928269 GCCGTGGCCGCCCGGACACACCTTGGCCGCGCTGATGCGGGGTCCCGCTCGCGCGCGC 3928210
QY 1463 aagttctcgcaattcaccgcgaacacagccatccagagagcgttcgagcgagagagcgc 1522
Db 3928209 GCCCGCGGCGCCCG 3928150
QY 1523 gtgcacatcatgcagcctgcagatcatgcaagggtgagtgagtgagcgagcgcttcacatc 1582
Db 3928149 GCCGATTCGGCGCTTGTGCGCCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3928090
QY 1583 ctgcctccgc 1642
Db 3928089 GTTGGCGCGCGCTCCCG 3928030
QY 1643 gagcgctcgagcgcaagcgagcgagcgcttcgatttcgcgacagcgctcgagcgcc 1702
Db 3928029 CAGGGTGGCGGCG 3927970
QY 1703 ttcgatttcgagcg 1762
Db 3927969 GTTGATGCG 3927910
QY 1763 acgagcgagagcg 1822
Db 3927909 AAGGCGGAGAGGCG 3927850
QY 1823 tcgagctcgaacagcgctcgatcgc 1850
Db 3927849 GCCGTTTCG 3927822

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 diagonal Road, Suite 500
; City: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMUG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109

; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-Fls
; US-08-232-463-14

Query Match 3.1%, Score 107.6; DB 1; Length 7218;
Best Local Similarity 3.5%; Pred. No. 1.4e-13;
Matches 14; Conservative 269; Mismatches 113; Indels 0; Gaps 0;

QY 103 gccgcgagcgagcgaacgaacgcaactgctcctcctcctcctcctcctcctcctcctcctc 162
Db 1041 GGGTGCAGGTGAGGAGGAGCTTGTGCGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1100
QY 163 ctggagtcgc 222
Db 1101 TT 1160
QY 223 tctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 282
Db 1161 TT 1220
QY 283 cccacactcaccgacgc 342
Db 1221 TT 1280
QY 343 cctcccccacacctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 402
Db 1281 TT 1340
QY 403 cctccatcagctagaccaccgaagcgcgcgcgcgccatgctcgcaagcgccgcgc 462
Db 1341 TT 1400
QY 463 gtcgacatgagcctccgc 498
Db 1401 TT 1436

RESULT 4
US-07-945-283-1/C
; Sequence 1, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EPO and LTR Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; City: Peoria
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P


```

Db      4185 CTCCTCCGGCTCTCCGGCTACGACCAAGCTTCCCCCGCCGCGGAC 4138

RESULT      5
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 441529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Query Match	2.7%;	Score 94;	DB 4;	Length 4411529;
Best Local Similarity	45.1%;	Pred. No. 1.4e-09;		
Matches 738;	Conservative 0;	Mismatches 873;	Indels 24;	Gaps 9

[illegible][illegible]

RESULT 6
US-09-103-840A-2


```

Db 3548 GGAGCCACCGCTAGAACCGCTCCGAGTCTGCGACTGATGTTGGACGATCCGCC 3489
OY 703 CGGTCGACGAGCGCG-----GGGTGAGCGCATCATCCGGAATCATCGGGA 752
Db 3488 GCGCGCGCCACCGGAGAACCACTCCGACGCTGCTGCGCGCGCCGCGGAATGCGGCC 3429
OY 753 GCAAGCGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 812
Db 3428 GCGCGCGCGCGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3369
OY 813 ACCCTGCAACCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 872
Db 3368 GCTGCTTCCACCGCGCGCGGGAATGCGCGCTCCGAGCTGAGCGCACCGCGCCAG 3310
OY 873 CGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 932
Db 3309 AGTACTTCCCGCTCCGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3250
OY 933 CTCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 992
Db 3249 CTCGCGACTAGAGCGCGCGCGCGCTCTGATGCTGCGCGCGCGCGCGCGCGCGCGCG 3190
OY 993 GCGCGCAAGCATCCACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1052
Db 3189 CCGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3130
OY 1053 CGCGAGTGC-----CGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1104
Db 3129 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3070
OY 1105 CGCAACAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1163
Db 3069 CGCGCGACTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3010
OY 1164 AGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1205
Db 3009 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2968

RESULT 10
US-08-458-240-1/c
: Sequence 1, Application US/08458240
: Patent No. 6143727
: GENERAL INFORMATION:
: APPLICANT: Roop, Dennis R.
: APPLICANT: Rothmeyer, Joseph A.
: APPLICANT: Greenhalgh, David A.
: TITLE OF INVENTION: SPECIFIC EXPRESSION VECTORS
: TITLE OF INVENTION: AND METHODS OF USE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LYON & LYON
: STREET: 611 West Sixth Street
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90017
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: WordPerfect (Version 5.1)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,240
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/146,930
: FILING DATE:
: APPLICATION NUMBER: 07/876,286
: FILING DATE: April 30, 1992
: APPLICATION NUMBER: No. 6143727 yet assigned (204/132)
: FILING DATE: October 29, 1993

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 204/152
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ. ID NO. 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6530 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-458-240-1

Query Match 2.5%; Score 88.4; DB 3; Length 6530;
Best Local Similarity 46.3%; Pred. No. 1.5e-09;
Matches 519; Conservative 0; Mismatches 581; Indels 22; Gaps 6;

OY 104 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 163
Db 4087 CCGTTCACACCTTCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4028
OY 164 TGGTTCGCGCGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 223
Db 4027 CGCGCTCCGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3968
OY 224 CCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 283
Db 3967 CAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3908
OY 284 CCGACTCCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 343
Db 3907 CCGTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3848
OY 344 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 403
Db 3847 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3788
OY 404 CTCGATCAGCTAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 463
Db 3787 CCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3728
OY 464 TCGGATGAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
Db 3727 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3669
OY 524 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582
Db 3668 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3609
OY 583 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
Db 3608 CCGGAGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3549
OY 643 GGTGAGCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 702
Db 3548 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3489
OY 703 CGGTCGACGAGCGCG-----GGGTGAGCGCATCATCCGGAATCATCGGGA 752
Db 3488 GCGCGCGCCACCGGAGAACCACTCCGACGCTGCTGCGCGCGCGCGCGGAATGCGGCC 3429
OY 753 GCAAGCGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 812
Db 3428 GCGCGCGCGCGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3369
OY 813 ACCCTGCAACCCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 872
Db 3368 GCTGCTTCCACCGCGCGCGGGAATGCGCGCTCCGAGCTGAGCGCACCGCGCCAG 3310

```

[illegible]

RESULT 11
 PCT-US93-03993-1/c
 : Sequence 1, Application PC/7US9303993
 : GENERAL INFORMATION:
 APPLICANT: Roop, Dennis R.
 APPLICANT: Rothmangel, Joseph A.
 APPLICANT: Greenhalgh, David A.
 TITLE OF INVENTION: CONSTITUTIVE AND INDUCIBLE EPIDERMAL
 TITLE OF INVENTION: VECTOR SYSTEMS
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fulbright & Jaworski
 STREET: 1301 McKinney, Suite 5100
 CITY: Houston
 STATE: Texas
 COUNTRY: U.S.A.
 ZIP: 77010-3095
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/03993
 FILING DATE: 19930428
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Paul, Thomas D.
 REGISTRATION NUMBER: 32,714
 REFERENCE/DOCKET NUMBER: D-5405
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 713/651-5325
 TELEFAX: 713/651-5246
 TELEX: 762829
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6530 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 PCT-US93-03993-1

[illegible]

Db 3069 CGCCGACGTGAGCGCGCCGCCCTCCGAGTAGCCACCTCCGACGTGAGCGCGCC 3010
QY 1164 agcgagcagagagagcgctccactctgtctgacgtctgtcgc 1205
Db 3009 CGCCCGCAGAGTAGCACCGCCGCCGAGTAGACCCACCGCCG 2968

RESULT 12

US-09-165-264-7/c
Sequence 7, Application US/09165264
Patent No. 6197510
GENERAL INFORMATION:
APPLICANT: Vinyagamoorthy, Thuralayah
TITLE OF INVENTION: Multi-Loci Genomic Analysis
FILE REFERENCE: 44747
CURRENT APPLICATION NUMBER: US/09/165,264
CURRENT FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 7
LENGTH: 320
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Primer sequence
US-09-165-264-7

Query Match 2.5%; Score 88.2; DB 4; Length 320;
Best Local Similarity 54.9%; Pred. No. 4.8e-10;
Matches 174; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 198 ctcttgagcgttctagagagctctctctctctctctctctctctctctctctcgc 257
Db 319 CTAGAGGATGCTAGCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 260
QY 258 cgcacacacgc 317
Db 259 CC 200
QY 318 ctctacactctatctgc 377
Db 199 CC 140
QY 378 ccccaaacactctctctctcaactcaactcaactcaactcaactcaactcaactcaactcaact 437
Db 139 CC 80
QY 438 cgcacatggtcgcgaagc 497
Db 79 CC 20
QY 498 tcacgagcgacgtctcc 514
Db 19 CCCCCCCCCCCCCCCCCC 3

RESULT 13

US-09-050-863-2/c
Sequence 2, Application US/09050863
Patent No. 6114111
GENERAL INFORMATION:
APPLICANT: Lao, Ying
APPLICANT: Hiang, Betty
APPLICANT: Pavan, Don
TITLE OF INVENTION: Mammalian Protein Interaction Cloning
TITLE OF INVENTION: System
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Teal, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050,863
FILING DATE: 30-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1969
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-09-050-863-2

Query Match 2.5%; Score 87.2; DB 3; Length 2580;
Best Local Similarity 47.8%; Pred. No. 1.8e-09;
Matches 378; Conservative 0; Mismatches 403; Indels 9; Gaps 4;

QY 205 cgttctagatgagctcgc 264
Db 1520 CTTTCTTCACAGTCGACGACACTCTCCCTGCTCTTTACAGACTCCCTCCCTGCTCT 1461
QY 265 accgcccccaactccttcccaactcctcagcagcagcagcagcagcagcagcagcagc 324
Db 1460 TTACAGCTGCTACCCGCGGCGCTTCACCTCTCTCCACCCCGGCTCCACTGCTCC 1401
QY 325 tctctatcttcgc 384
Db 1400 TCGACCCCGGCTTCAGTCTCTCTGACCCGCGGCTTCACCTCTCTGCTCTGCTCTCT 1341
QY 385 ----ccactctctctcactcactcactcactcactcactcactcactcactcactcactcact 440
Db 1340 TGTCTGCTGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1281
QY 441 ccatggtcgcgaagc 500
Db 1280 CCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1222
QY 501 cggagcagctctcgcagc 560
Db 1221 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1162
QY 561 cgtctctcgcagc 620
Db 1161 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1104
QY 621 cgttccagacacacgc 680
Db 1103 TGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1044
QY 681 cgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 740
Db 1043 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 984
QY 741 acatctcggagcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 800
Db 983 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 924
QY 801 gcgagatcatcaacccctcgaaccccgagcgtcggtctgctctgagctcgcgcgcgcgcgc 860

Db 1566 CTTTTCACAGTCACAGACTCTCCCTGCTCTTTCACAGACTTCCCTCCGCTCT 1507
OY 265 accgcccccaactccttcccaactcaltgacacgcacacgcctccctcctc 324
Db 1506 TTTCAAGTCTTACCCGCGGCTCCACTCTCTGACACCCGCGCTCCACTACCTCC 1447
OY 325 tccattatgcgcgcgtgcctccccaactccctccgtcctcctcctcctcccaaa 384
Db 1446 TCGACCCCGGCTCCACTCTCTGACACCCGCGCTCCACTCTCTCTGCTCCCTCC 1387
OY 385 ----ccactcctcccaactcactcactcactcactcactcactcactcactc 440
Db 1386 TGTCT 1327
OY 441 ccatggtcgcgaagc 500
Db 1326 CCT 1268
OY 501 cggcggaactctcgc 560
Db 1267 CTCTGCT 1208
OY 561 cgtcctcgc 620
Db 1207 CTCTGCT 1150
OY 621 cgttcagacacacgc 680
Db 1149 TCCCT 1090
OY 681 cgcagcg 740
Db 1089 TGTCT 1030
OY 741 aactatcgggagcagcg 800
Db 1029 TCT 970
OY 801 gcgagatatacaccctcgaaccccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 860
Db 969 CCT 910
OY 861 cccctcgc 920
Db 909 TCT 850
OY 921 tccctgcacgc--gctcgc 978
Db 849 TCT 790
OY 979 gcttcggac 988
Db 789 CCT 780

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 22, 2002, 14:34:04 ; Search time 2858.62 Seconds
(without alignments)
13194.367 Million cell updates/sec

Title: US-09-265-585c-95

Perfect score: 3510

Sequence: 1 ctgctagctcagcctactca.....tagccatcaagcaagtatla 3510

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gp_estl:*
11: gp_est2:*
12: gp_hic:*
13: gp_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_huv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_tod:*
20: em_gss_vrl:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	498.4	14.2	715	10	AU093320 AU093320
C 2	466.2	13.3	586	10	BE359444 BE359444
C 3	378.6	10.8	665	10	BE366082 BE366082
C 4	201.8	5.7	528	11	BE154957 BE154957
C 5	194.6	5.5	431	10	BE209394 BE209394
C 6	193.2	5.5	504	10	AM004485 AM004485
C 7	175	5.0	1015	11	BI416573 BI416573
C 8	170.2	4.8	1223	11	BI416573 BI416573
C 9	165	4.7	1197	11	BE265162 BE265162
C 10	164	4.7	484	11	BI311576 BI311576
C 11	163.4	4.7	728	11	BE267681 BE267681
C 12	162.4	4.6	1299	11	BE267681 BE267681

C 13	161.8	4.6	1188	10	BE455091 BE455091
C 14	161.6	4.6	931	11	BI416575 BI416575
C 15	161	4.6	1050	11	BE267685 BE267685
C 16	159.8	4.6	893	13	AZ199082 AZ199082
C 17	155.8	4.4	937	11	BI416476 BI416476
C 18	155.4	4.4	1039	11	BE259783 BE259783
C 19	155	4.4	1224	11	BI416537 BI416537
C 20	154.8	4.4	1080	10	BE455154 BE455154
C 21	154.6	4.4	1042	11	BI416535 BI416535
C 22	152	4.3	1135	11	BE265506 BE265506
C 23	152	4.3	1161	11	BE265588 BE265588
C 24	150.4	4.3	1195	11	BI416506 BI416506
C 25	149.4	4.3	1088	11	BE265580 BE265580
C 26	149.2	4.3	1277	11	BG447302 BG447302
C 27	147.8	4.2	1107	11	BI416477 BI416477
C 28	147.8	4.2	1342	10	BE455093 BE455093
C 29	147	4.2	924	10	BE455162 BE455162
C 30	146	4.2	1035	11	BE254665 BE254665
C 31	145.4	4.1	1006	10	BE034822 BE034822
C 32	145.4	4.1	1116	11	BE256617 BE256617
C 33	145.2	4.1	1248	10	BE455160 BE455160
C 34	144.6	4.1	1325	11	BE265543 BE265543
C 35	143.8	4.1	563	11	BE588097 BE588097
C 36	143.6	4.1	952	11	BG441341 BG441341
C 37	143	4.1	883	13	AZ186646 AZ186646
C 38	143	4.1	1028	11	BG809907 BG809907
C 39	142.6	4.1	700	10	AL506960 AL506960
C 40	142.6	4.1	1080	13	AO893056 AO893056
C 41	141.8	4.0	1166	11	BE256751 BE256751
C 42	141.6	4.0	999	11	BI416539 BI416539
C 43	141.4	4.0	1195	11	BE256755 BE256755
C 44	141	4.0	1328	10	BE455227 BE455227
C 45	140.4	4.0	551	10	AI770444 AI770444

ALIGNMENTS

RESULT 1
LOCUS AU093320 715 bp mRNA
DEFINITION AU093320 Rice callus Oryza sativa cDNA clone C61354, mRNA sequence.
ACCESSION AU093320
VERSION AU093320.1 GI:8594782
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 715)
Sasaki, T. and Yamamoto, K.
Rice cDNA from callus (2000)
Unpublished (2000)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abrr.affrc.go.jp, URL: http://irgp.dna.affrc.go.jp/PROJECT/"RGP".
C61354_87.

FEATURES

source
location/Qualifiers
1..715
/organism="Oryza sativa"
/strain="Cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone="C61354"
/clone_1bp="Rice callus"
/note="Vector: pBluescript II SK+; Site 1: SalI; Site 2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site

RESULT	3
BE366082	
LOCUS	
DEFINITION	BE366082 665 bp mRNA EST 20-JUL-2000
ACCESSION	PI1_30_C11.g1_A002 Pathogen induced 1 (PI1) Sorghum bicolor cDNA,
VERSION	BE366082
KEYWORDS	mRNA sequence.
SOURCE	BE366082.1 GI:9307639
ORGANISM	EST.
	sorghum.
	Sorghum bicolor
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
	clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE	1 (bases 1 to 665)
AUTHORS	Cordonnier-Pratt, M.-M., Gingle, A., Dean, R., Sudman, M. and Pratt
	, L.H.
TITLE	An EST database from Sorghum: pathogen-induced plants
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cordonnier-Pratt MM

Email: mmpmratt@uga.edu
Sequences have been trimmed to exclude POLYA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: Polymix
High quality sequence start: 16
High quality sequence stop: 665
POLYA-No.

FEATURES	Location/Qualifiers
source	1. .665

1. .665
/organism="Sorghum bicolor"
/db_xref="taxon:4556"
/clone_lib="Pathogen induced 1 (PI)"
/note="Organ: Anthracnose-infected leaves from
two-week-old sorghum plants 48 hr after inoculation;
Vector: plasmidicript II from Lambda Zap II; Site,1 : XhoI;
Site,2 : EcoRI; Two-week-old sorghum plants (BTX 623
cultivar) were infected with pathogen (isolate FRM21 of
Colletotrichum graminicola, which is a sorghum isolate).
RNA was prepared from infected leaves harvested from 45
seedlings 48 hours after inoculation. Note: young
seedlings (2 weeks old) exhibit juvenile resistant
reaction, which is an incompatible interaction. As they
grow older (4 weeks or older), plants were made susceptible
to anthracnose disease. The library was made from poly-A
RNA in the cloning vector lambda Zap II. Clones to be
sequenced were prepared by mass excision. WARNING: While
most or all ESTs are expected to derive from the host
plant, no effort was made to eliminate ESTs deriving from
the pathogen."

BASE COUNT	123 a	197 c	187 g	158 t
ORIGIN				

Query Match	10.88;	Score 378.6;	DB 10;	Length 665;
Best Local Similarity	96.58;	Pred. No. 3.5e-47;		
Matches 387; Conservative	0;	Mismatches 14;	Indels 0;	Gaps 0;

QY		cttccctgagcgcgttcgtgtgaggacatccactactactcgcgctgtltcgactcgctga	2815
Db	1	CTTCCCTGAGCGCCTTCGTGGAGGCCATCACTACTCTGCGCGCTGTTCACCTCGCTGA	60

OY	2816	cgcgagctacgcgcgaagacagcccgagcgcgcactcgttgagcaacgactcgttcgcg	2875
Db	61	CGCAGACTACGCGGAGGACAGCCCCAGCGGCACGTCTGGAGCACAGCACTGCTGTCCGC	120

[illegible]

D _b	121	GGAGATGCGCAACGTGCTGGCCGTCGGCAGGCCGCCGCCGCAGCCGGGAGACTGAATTGGG	180
Q _y	2956	cagctgtgcgcgagaagtcttggccgafttcggygttcgcgcgcgcctcgtctgcgcggaagcgc	2999
D _b	181	CAGTGTGGGCGCAGAAAGTCTCGCGCATGCCGGTTTCCGCGCGCTCTCTCTGCCCCGAGCGCC	240
Q _y	2996	cgcgcgcgaagcgcgcctctgtctctgcgcgcacgtttcccctccgaagcgtagaacgttgttga	3055
D _b	241	CGCGCGCGCAGGCGTCGCTGCTGCTCGGAATGTTCCCTTCGAGAGSSTACACCGTGSTGA	300
Q _y	3056	ggagaaacgcgcgccttgaagcttcgggttgaaagaacctctgccttcaaccgcgtcggccty	3115
D _b	301	GGAAGAAGCGCGCGCTCTCAACGTCGCGCTGGAGAAGCACTCTGCTCTCTCACCGCGCTCGGCA TG	360
Q _y	3116	ggcgcgccatccagttgcgcgcgcgcgttcgacttgaataaacctct	3156
D _b	361	GCGCCCATCCAGATGCCCGCGCGCTGACCCTGTAATAGAGCTGT	401

RESULT	4
BGI54957	
LOCUS	BGI54957
DEFINITION	BGI54957 528 bp EST
ACCSSION	sab28808.y1 Gm-cl026 glycine max cDNA clone GENOME SYSTEMS CLONE
VERSION	ID: Gm-cl026-2775 5 similar to TR:Q96304 Q96304 SCAECROW. ; mRNA sequence.
KEYWORDS	BGI54957 BGI54957 GI:12688621
SOURCE	EST.
ORGANISM	soybean.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots: Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE
AUTHORS

AUTHORS
Shoemaker, R., Kaim, P., Voklin, L., Expediting, J., Corryell, V., Khanna, A., Bolla, B., Merritt, M., Hillier, L., Kricaba, T., Matlin, J., Beck, C., Wylie, T., Underwood, K., Stepte, B., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE
Public Soybean EST Project

JOURNAL
Unpublished (1999)

COMMENT:

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info.genomesystems.com web site: www.genomesystems.com
High quality sequence stop 455.

FEATURES

source

```

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1026-2775"
/clone_lib="Gm-c1026"
/lssue_type="Senescing leaves, mature plants, greenhouse
grown."
/lab_host="DH10B"
/notice="Vector: pUT19-Pac (Pharmacia); Site_1: EcoRI;
Site_2: HindIII; This cDNA library was constructed from
mRNA isolated from senescing leave tissue of mature
greenhouse grown plants. Complementary DNA was synthesized
from mRNA using a 3' anchored poly(dT) primer. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by digestion with EcoRI and HindIII. The cDNA
fragments were directionally cloned into the EcoRI-HindIII

```

0Y 2876 ggagatccgcgaacgtgctgctggcgccgcgacgcgacgtcaagtctcg 2935

RESULT	6	AM004485	504 bp	mRNA	EST	08-SEP-1999
LOCUS		701931804	A. thaliana	mixed source	Arabidopsis thaliana	cDNA clone
DEFINITION		701931804	mRNA sequence.			
ACCESSION		AM004485				
VERSION		AM004485.1	GI:5851514			
KEYWORDS		EST.				
SOURCE		thale cress.				
ORGANISM		Arabidopsis thaliana				
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS		1 (bases 1 to 504)				
		Chen, J., Montanya, M., Chan, E., Mooney, M., Carreon, B., Gilliland, D., Wang, K., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzowska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policy, J., Suzuki, G., Argentine, C., Shah, S., Nobriga, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.				
TITLE		Arabidopsis thaliana Gene Expression Microarray				
JOURNAL		Unpublished (1999)				
COMMENT		Contact: David Smoller, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 4633 World Parkway Circle, St. Louis, MO 63134, USA Tel: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com.				
FEATURES		Location/Qualifiers				
source		1..504 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="701931804" /clone_1id="A. thaliana, mixed source" /note="This sequence was obtained from a clone generated with a PCR product of the target gene."				
BASE COUNT		137 a 93 c 130 g 144 t				
ORIGIN						
Query Match		5.5%; Score 193.2; DB 10; Length 504;				
Best Local Similarity		66.1%; Pred. No. 1.5e-19;				
Matches		279; Conservative 0; Mismatches 143; Indels 0; Gaps 0;				
2700		agcgctgcccccaagctgtgtacaaatgltgtgaagcagcagcttgagccactcgagctccctc				2759
Db		73 AGATTAGCTCTCTAAAGTTGTGTGACACTAGTCGACAAAGATTGAGCGACAGCTGTCTTTC				132
2760		ctggcgccctcttggaagccatccactactactgsgcgttgactgactcgctgagccg				2819
Db		133 TTAGGAACATTTGTGAAGCAATATATTACTGTGACTCTTTGACTCACTGGAGCA				192
2820		agctacgcgcagaggaagcccccagcgagcagctctgtgagcagcaagctctgtcgcgga				2879
Db		193 AGCTACGGCGGAAGAGTAGAAGAGAGACATGTCGTGGAACAGCAGACTATTATCGAAGAG				252
2880		atccgcaacgtgtctgtgcgttggcgggcgccgcgcgcacccgcgcagcctcaagttcgca				2939
Db		253 ATACGAATGTATTGTAGCGTTGAGGACATCGAAGACGGGTGAAGTTGAGAGC				312
2940		tggcgccgagagctggcgcagctcggggttcgcggcgccctgctcgccgcgcgcgcgcg				2999
Db		313 TGGAGGGAGAAATCAACAAATGTGGGTTTAAAGGTATATCTTTAGCGGAATCCACT				372
3000		gcgcgcgcgtccctctgtctgcgcagctgtccctcccgacgcgcgttgcgcgttgcgcgc				3059
Db		373 ACACAGGAGACTCTACTCTGTGGGATGTCTTCCTTGCGATGTGTTACACTTTGGTTATGAT				432
3060		aacgcgcgttgaagctcgggtgtgaagagcactctctgtctcaaccgcgttcgcgcgcgc				3119

[illegible]

[illegible]

RESULT	9
BI416470/c	
LOCUS	BI416470
DEFINITION	1197 bp mRNA
	EST
	15-AUG-2001
	hasp001xa01f Heterobasidium annosum - Scots pine infection stage

ACCESSION	BI416470
VERSION	BI416470.1
KEYWORDS	GI:15187493
	EST.

SOURCE	ORGANISM	REFERENCE
Heterodactylodon anomus	Pinus sylvestris mixed EST library.	1 (bases 1 to 1197)
Heterodactylodon anomus	Pinus sylvestris mixed EST library	1 (bases 1 to 1197)
Eukaryota, mixed EST libraries.		
Asieghu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.		
Expressed sequence tags of randomly selected cDNA clones from the		

JOURNAL
COMMENT
Unpublished (2001)
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026 S-750 07 Uppsala

Tel.: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Astlebu@nykopat.slu.se
Seq primer: T7 primer
Location: *dnalst1::f1arc*

```
/organism="Heterobasidion annosum/Pinus sylvestris mixe
EST library"
/db.xref="taxon:169015"
/clone="hasp001xa01f"
```

```
/dev_stage="Seedling roots of scots pine were infected:  
6 days with H. annosum"  
/note="Vector: pT-Adv1; Site_1: EcoRI. The subtractive  
hybridization cDNA library was constructed from scots
```

BASE COUNT ORIGIN	20	a	28	c	1036	g	28	t	85	others
----------------------	----	---	----	---	------	---	----	---	----	--------

Matches	462;	Conservative	0;	Mismatches	545;	Indels	2;	Gaps	
---------	------	--------------	----	------------	------	--------	----	------	--

[illegible]

OY 461 ggcctcgacatggaactccgcgcgcgcgcgcgcacgttaacggcgaactctcagaatc 520
 | | | | | | | | | | | | | | |
Db 822 CANNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCNCCCCCCCCCCCCCCC 763

581 ctgcgcgcgtgccaccacagctccaccagctgccccgcgttcagacaccagcgccg 640

Db	702	CNNCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	643
Qy	641	gaggtgagctgtccgcgcacccggcccccgcgcgcacgcgagcgagggcgagcagca	700
Db	642	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCNCCCC	583

D_b 582 CCCCCCCCNCNCCGCCACGACCAGCAGACCCCCCNCCNNCCCCCCCCCACACC 523
Q_y 761 ggcgacgcggtctccatcacgcagtcattcacaacgltccggagatcatlcaaccctgc 820

	821	aaccgcgactcgtctcgctctgagatccgcgctccctccctcgaacgcgacccg	880
Oy			
Dd	462	ccccccccccccccccccnncccccccccccccccccccccccnnccccnnnc	403

Dc
941 ggcgctcccgagggatgtacgtcttcctcccgcgaacgagtcttcgagaagagccgccac 1000

Oy
402 ccccccccccccccgcaccgcccccccccccccccccccccccccccccccccc - CC 345

DY 1001 gagcatccacgcgcttcagcagcaacagcagcgaaacatccgcgcgcgactcg 1066
|||
Db 284 NCGNNCCGCCCCCNCACAAACCNCNCCCCCCTGCCCNCCGNCNCCGTTCNCCCTCNCNC 225

224 CCCCTCANNCCGGCNTNCCAGAGCCCGCGTGGCGCCAGGCCCCCCCG 176

LOCUS	BF265162	484 bp	mRNA	EST	09-MAR-2001
BF265162					

Db	327	AACATGCGTAAACGTCATTTGGTGGCGCGCTGTCAGTCTTGTTCGCTACTTACGAGCTGCTCG	386
Oy	3218	gtttggtc 3225	
Db	387	GTGGCATC 394	
RESULT	11		
LOCUS	BI311576		
DEFINITION	BI311576 728 bp mRNA EST5133366 GBSD Medicago truncatula cDNA clone pGBSD13C12 5' end,		
ACCESSION	BI311576		
VERSION	BI311576.1	GI:14985903	
KEYWORDS	EST.		
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;		
	Medicago.		
REFERENCE	1 (bases 1 to 728)		
AUTHORS	Grusak, M.A., Sanac, D.A., Town, C.D., Van Aken, S., Uterback, T., Cho, J., and Fraser, C.M.		
TITLE	ESTs from developing reproductive tissues of Medicago truncatula		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Michael A. Grusak USDA/ARS Children's Nutrition Research Center Baylor College of Medicine 1100 Bates Street, Houston, TX 77030-2600, USA Tel: 713-798-7044 Fax: 713-798-7078 Email: mgrusak@bcm.tmc.edu B398431e		
FEATURES	TIGR sequence name: MTPAV19TK More information is available at: www.medicago.org Seq primer: SKmd (CTA GAA CTA gtc gat CC).		
SOURCE	Location/Qualifiers 1..728 /organism="Medicago truncatula" /cultivar="A17" /db_xref="taxon:3880" /clone="pGBSD13C12" /clone_lib="GBSD" /russue_type="Immature seeds" /dev_stage="Immature seeds, 11 to 19 days after pollination" /note="Vector: pluscript SK-; Site_1: EcoRI; Site_2: XhoI; Immature seeds, collected from pods ranging in age from 11 to 19 days after pollination, were harvested from greenhouse-grown plants. Seeds were removed and separated from pod walls and were immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the unzip xr vector from stragene and packaged using Gibapack III gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in XL0IR cells."		
BASE COUNT	204 a 123 c 168 g 233 t		
ORIGIN			
Query Match	4.7%; Score 163.4; DB 11; Length 728;		
Best Local Similarity	64.3%; Pred. No. 4.1e-15;		
Matches 245; Conservative	0; Mismatches 136; Indels 0; Gaps 0;		
Oy	2741	gagcaccgcgggctccctccgcgcgcgttgtagagccatccacactactcggcgcct 2800	
Db	1	GAGCAATGCAAGGCTCATTTCTTGGGAAGGTTTGGAAGCAATACATTACTCTACGACTT 60	
Oy	2801	gttcgaccctgcttgacgagagctacgcgagaggaacagcccccagcgacgctctgtagca 2860	

[illegible]

RESULT	12
LOCUS	BF267681
DEFINITION	BF267681 1299 bp mRNA EST 09-MAR-2001 HVCSE0018J13f Hordeum vulgare seedling green leaf EST library HVCN00004 (Erysiphe infected & control) Hordeum vulgare cDNA clone HV_CSE0018J13f, mRNA sequence.

ACCESSION	BF26/681
VERSION	BF267681.1
KEYWORDS	GI:1198676 EST.

ORGANISM
Hordeum vulgare
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae, Pooidae
Triticeae: Hordeum.
1 (bases 1 to 1299).
Wing, R., Close, T. J., Kleinof, A., Wise, R., Begum, D., Frisch, D., Yu,
X., Anderson, H., Dale, J., Henry, D., Kenndle, S., Palmer, M., Rambo,
T., Sasaki, C., Schwartzbeck, J., Simons, J., Choi, D. W., Main, D. and
Wood, T.

TITLE	Development of a genetically and physically anchored EST resource for barley genomics
JOURNAL	Unpublished (2000)
COMMENT	Contact: Wing RA

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864.656.2800

Email: twing@clmson.edu
Seq primer: AATTACCCCTCACTAAAGG
High quality sequence start: 321
High quality sequence stop: 1283

FEATURES	Location/Qualifiers
source	1. .1299

```

/organism="Hordelum vulgare"
/cultivar="C116155 (Mlat3)"
/db_xref="taxon:4513"
/clone="HV_CEA001813jf"
/clone_lib="Hordelum vulgare seedling green leaf EST
library HVCDA0004 (Erysiphe Infected & control)"
/tissue_type="seedling green leaf"

```

```

/note=Vector: lambdaZAP, Site_1: EcoRI, Site_2: XhoI. For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/. To
order a clone see http://www.genome.clemson.edu/orders"
BASE COUNT      0 a      827 c      82 g      0 t      390 others
ORIGIN

```

Query Match	4.6%;	Score 162.4;	DB 11;	Length 1299;
Best Local Similarity	42.9%;	Pred. No. 6.1e-15;		
Matches 441; Conservative	0;	Mismatches 583;	Indels 4;	Gaps 1.

[illegible]

QY	1064	aag	cccc	1071
Db	1265	cccc	cccc	1272

RESULT 13	BE455091/c	LOCUS	DEFINITION
BE455091	1188 bp	mRNA	EST
HYSMENH0096C14f	Hordeum vulgare	5-45 DAP spike	EST library
HVCDNA0009 (5 to 45 DAP)	Hordeum vulgare	clone	HYSMENH0096C14f.
mRNA sequence.			

